HOW PEOPLE INTERPRET AND REACT TO EVERYDAY AUTOMATION ISSUES

A Thesis
Presented to
The Academic Faculty

by

Kimberly C. Preusse

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology
August 2016

HOW PEOPLE INTERPRET AND REACT TO EVERYDAY AUTOMATION ISSUES

Approved by:

Dr. Wendy A. Rogers, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Francis T. Durso
School of Psychology
Georgia Institute of Technology

Dr. Rick P. Thomas
School of Psychology
Georgia Institute of Technology

Date Approved: May 5 ,2016

ACKNOWLEDGEMENTS

I am very grateful for Dr. Wendy Rogers for her mentoring, support and guidance on this project. I also wish to thank my committee members, Dr. Frank Durso and Dr. Rick Thomas, for their valuable input in this endeavor. I would also like to acknowledge the members of the Human Factors and Aging Laboratory, especially Dr. Tracy Mitzner, Dr. Akanksha Prakash, Laura Barg-Walkow, and Sean McGlynn for their help with material development. Similarly, I would like to thank Sadaf Kazi and Ashley Ferguson for their guidance with data analysis as well as Xi Zhang and Kerri Read for their hard work on assisting with data analysis. I also thank my family and friends for the unwavering support. This research was supported in part by a grant from the National Institutes of Health (National Institute on Aging) Grant P01 AG17211 under the auspices of the Center for Research and Education on Aging and Technology Enhancement (CREATE).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Automation can be defined as “any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by a machine” (Moray, Inagaki, & Itoh, 2000, p.44). Although automation is ubiquitous and offers benefits to using it, automation can err. The human-automation interaction literature has shown that people are less likely to use automation when the automation makes frequent errors (e.g., Wickens & Dixon, 2007).

However, the current human-automation interaction literature lacks a sufficient understanding of what people do when automation does make an error. Research in the human-automation literature (e.g., Dzindolet, Peterson, Pmranky, Pierce & Beck, 2003; Madhavan, Weigmann, & Lacson, 2006) and person perception literature (e.g., Kelley, 1967; Pop, 2013) suggest that users might rely on a variety of cues to know that an automation issue has occurred. Although these studies assume or suggest people differentiate between qualitatively different types of errors (e.g., errors with different causes), research has not examined how this interpretation occurs. It remains unknown with what level of detail experienced everyday users interpret automation issues.

Little is also known about what happens after users interpret an automation issue. Most of the studies in the human-automation have constraint the possible responses of participants to either rely on the automation more or rely on it less (e.g., Muir & Moray, 1990, Bisantz & Seong, 2011). Thus, an open question is what users actually do in their everyday lives when those constraints are lifted. A related open question is how do users decide which response to take.

This thesis qualitatively answered gaps in the literature regarding how people interpret and react to automation issues. Thirty experienced everyday automation (activity tracker) users each completed a Threat-Strategy Interview and a Scenario-Based Interview. These methods elicited rich, detailed responses of cues to automation issues, the interpretation of issues, reasons for responses, and responses.

Findings support the generalizability of the cue to issues found in experiments to everyday automation. Additionally, the results revealed other cues not currently found in the human-automation literature, such as measurement comparison. Findings do not suggest that cues inherently result in causal interpretations of the issue. Rather, users might interpret an automation error generally or may just have knowledge of the location of the error. Additionally, users do not necessarily have only one interpretation of an automation issue.

With regard to responding to an automation issue, data reveal the roles many reasons have in deciding what action to take. These reasons included: various types of knowledge; the importance of the issue for the user's purpose of using the technology; the ease of implementation of the action; the lack of feasible alternative actions; the extent of the issue, and the situational consistency of the issue. Further, the present study also documents reaction strategies related to fixing the issue on their own, getting help, and changing the user's behavior. These strategies had previously not been found in the human-automation literature on errors.

The data from the two interviews were integrated to create conceptual frameworks of how automation issues are interpreted and reacted to. Understanding how individuals interpret and react to automation errors is the first step in improving how individuals handle automation errors. Indeed, it will become increasingly important to guide successful responses to automation errors as automation usage increases.

CHAPTER 1

INTRODUCTION

Automation has the potential to make life easier by carrying out tasks effectively and freeing humans from time consuming activities (Parasuraman & Riley, 1997; Wickens & Dixon, 2007). Indeed, one popular definition of automation is “any sensing, detection, information-processing, decision-making, or control action that could be performed by humans but is actually performed by machine” (Moray, Inagaki, & Itoh, 2000, p.44 p. 44). However, automation is imperfect, and people must react to these imperfections. Yet, an in-depth assessment of how users interpret and react to automation imperfections or errors has yet to be conducted. This type of an assessment could guide error response training (i.e., troubleshooting). The current study examined attended automation error interpretation and reaction for a popular, health-related, everyday technology: wearable activity trackers.

An example of a wearable activity tracker is the Fitbit One (Figure 1). The Fitbit One uses automation to detect steps walked, calculate calories burned, sense time asleep and sleep quality, and sense stairs-climbed. However, wearable tracking devices like the Fitbit One are not completely accurate. Indeed, the prior iteration of Fitbit One was found to have an error rating of caloric expenditure of 10.1% (Lee, Kim, & Welke, 2014). In other words, the Fitbit One automation provided readings that were inaccurate by 10%. The automation mistake could be for a variety of reasons, such as inaccurately sensing steps walked. Mistakes such as inaccurately sensing steps, inaccurately calculating calories, and any other imperfection related to something the user does not have to do, can be considered an automation error. In addition, sometimes, the user may perceive an error occurred, even if one did not. For example, a tired user may perceive the sleep tracker as inaccurate if it said the user slept well, even if the user did indeed sleep well. Thus, a *perceived or attended to* automation issue can be defined as an instance in which the automation's outcome deviates from what the user expected to be correct.

It is currently unknown how experienced everyday automation users interpret automation issues (i.e., errors) and what strategies they might use to handle the issue. However, understanding how people trouble-shooting automation issues could substantial practical impacts. In particular, such an understanding would provide guidance to designers; training and help materials; and potentially other fields in marketing and deployment. For example, users might seek out particular cues to understand what is the issue. Designers may want to make those cues readily available in case of an automation issue. Marketing and deployment stakeholders might want to minimalize interpretations that a device limitation (i.e., something the device was not made to be able to do) are understood and are not classified as automation errors (e.g., to limit disappointment with the product). Similarly, if users do not understand what the issue is, they may not be able to find the issue in a help guide but may still need to be able to fix the issue. Indeed, one strategy for coping with an automation issue, especially if the user cannot easily fix it, may be to stop using it all together. Through understanding a user's reasons for selecting a strategy, training and help materials might make certain reasons (i.e., reasons associated with the optimal strategy) salient. For instance, if users face a complex issue where the optimal solution is to call a company or hotline for help, users might only do so if they had a reason to do so (i.e., the strategy was easy). To make the strategy easier, manuals or FAQs could include the contact information in the resource. In short, exploring how people troubleshoot automation issues is a first step to providing better human factors guidance for everyday automation users.

The Role of Knowledge in Interacting with Imperfect Automation

A rich human-automation interaction literature has incorporated automation issues, particularly automation errors, and recognized the importance of a user's knowledge of the automation capabilities (e.g., Riley, 1996; Sanchez, 2009). Such knowledge may include when, how, where, and for what the automation is reliable. In general, if a person knows an automation is not very reliable then he or she uses the automation less (Wickens & Dixon, 2007). For everyday technologies, a person's knowledge is largely built from experience. Compared to novice users, experienced users should better understand the automation because of more interaction with it (Norman, 2002; Van der Veer & Melguizo, 2003). Through interaction a user develops a device mental model for the automation (Norman 2002; Van der Veer & Melguizo, 2003).

A person's knowledge permeates through much of their daily life. For instance, general knowledge gained through life experience might help people attend to salient cues. Or, general knowledge might let people know that help for problems (e.g., automation issues) are typically available. Knowledge about the specific automation and causal knowledge help create a device mental model. These types of knowledge influence the very interaction with the automation (e.g., a person's decision to use the automation might be based off of knowledge about what task the automation is supposed to assist with). If a user knows enough about the automation, the user might even be able to understand the cause and react optimally. Thus, knowledge is especially important to keep in mind throughout the whole troubleshooting process of automation issue interpretation and automation issue reaction.

Issue Interpretation

A user must attend to an automation issue to decide to react to it. Thus, the first stage in troubleshooting is issue interpretation. Issue interpretation can further be divided into two parts:

how the user determines an issue occurred and what the user thinks is the issue. These two parts are issue interpretation are inherently linked.

Cues to Issues

Not all automation issues are necessarily automation errors. For instance, if a user understands the issue to actually be a device limitation, then the user may not consider it an automation error (Consolvo et al., 2008). It is also possible for an automation error to occur without the awareness of the user. This could occur if the user either does not have access to signs that there is an error or if the user is unaware that those signs are indicative of an error. Mental models can enable the interpretation of those signs as indicative of automation issues.

Device mental models

A device mental model is an individual's theory about how the automation works (Norman 2002; Van der Veer & Melguizo, 2003). Device mental models are conceptual models that mentally simulate the device's operation and are formed through interacting with the system (Norman, 2002). These models help users to try out alternative ways of using the technology, to utilize their past knowledge of the technology in future uses, and to react competently to unforeseen events such as automation errors (Van der Veer & Melguizo, 2003).

Device mental models form in part from a human tendency to come up with explanations (Norman, 2002). Therefore, device mental models may differ based on the situations that need explaining and may change over time as new experiences are added to the model (Van der Veer & Melguizo, 2003). Additionally, two users may interpret automation errors with different explanations because what constituents a satisfactory explanation varies from person to person (Craik, 1943). Lastly, because device mental models are formed from experiences and to differing satisfactory levels, the user's device mental model may not be the same as the designer’s model. Thus the user’s device mental model may not necessarily be representative of how the device actually works.

Theories in human-automation interaction, in part, have been extended from the human-human interaction literature (e.g., trust; Muir, 1994). The human equivalent of a device mental model in interpreting the behaviors of people is person perception. Thus, some of the cues that are used when interpreting human behavior may also be used in interpreting automation “behaviors” like errors.

<u>Cues to behavior in the person perception literature</u>

Cues presented along with the issue may help users determine how to accommodate or assimilate automation errors into their device mental model. When drawing from the attribution theories of the person perception literature (Kelley & Michela, 1980) it is important to note that automation errors differ from human behaviors in two important ways. First, observed human behaviors have more certainty in that they actually occurred, whereas a user may not always be certain if an error occurred. The second difference is that human behavior is broader whereas automation issues should be considered a *type* of “behavior.” That is, one “behavior” of the automation could be to work properly, whereas and automation issue would be a different type of behavior (e.g., working improperly). Human behavior might be praise-worthy, cringe-worthy, or anywhere in between, but for automation, I focus only on issues. Nonetheless, models from person perception are helpful in considering cues to automation issues.

It may be that automation issue interpretation parallels the human-behavior interpretation process. In extrapolating from Kelley and Michela (1980), device mental models (e.g., information and beliefs) help a user interpret what the issue is. What the issue is, in turn, guides the users reactions in responding (e.g., behavior toward) to the issue. Although the consequences of how the individual's behavior is interpreted (i.e., the automation error) typically have been dependent variables (e.g., decreased work motivation), in the case of automation errors, consequences may also be reasons for certain reactions. This might particularly be the case for the affective consequences of error interpretation (e.g., frustration encouraging a particular reaction). Lastly, experience understanding an automation issue might allow the user to know what that error is and what to do should that error occur again (i.e., expectancy in attributional theories).

The person perception literature also provides specific insights into the cues to automation issues. Accordingly, the causal attribution of the automation errors may depend on the consistency, consensus, and distinctiveness of the automation's behavior or functioning (Kelley, 1967; Orvis, Cunningham, & Kelley, 1975; Pop, 2013). With regard to consistency, the user might question how frequently the automation makes a particular error. In other words, the frequency with which the issue occurs is part of the extent of the error. As for consensus, the user might question if other, similar automation, makes the same particular error. In this way, consensus acts like a reference point, wherein the user can compare readings from two different devices. For situational consistency, also known as distinctiveness, a person might ask if the technology always errs around a particular stimulus. In this way, situational consistency is a large part of the context in which an issue might occur. These three cues of consistency (e.g., frequency), consensus (e.g., measurement comparison), and distinctiveness (e.g., context) might help form connections between an automation issue and the issue's cause. Individuals can then use these associations to revise their device mental models. In turn, the revisions to their device mental models will help them to interpret and react to future automation issues (i.e., through the user gaining knowledge).

Cues to errors in the automation literature

The typical design of human-automation interaction studies have made it difficult to understand the roles of cues to automation issues and most have only focused on automation errors. For example, reliability may be set to 70% for an experiment, with the 30% of errors occurring randomly (e.g., Riley, 1996). Participants are typically notified that an error occurred, but are typically not notified of the error's context. However, outside of the experimental setting, errors do not usually occur randomly and do occur in context. Thus, experience with random errors may not facilitate the development of understanding cues predictive of where or when to expect an error. The opposite may be true for experience with an everyday technology, where users may learn what signs are indicative of error.

Consistency-related cues to errors

Nonetheless, there have been some studies that infer cues to errors from the reactions of participants. In Itoh, Abe, and Tanaka (1999), the cue of distribution over time resulted in different usage patterns. After some experience using the system, if automation's errors occurred immediately following one another and then ceased, people returned to using automation faster than if the same number of errors occurred spaced out over time. The data may suggest that the differing reactions between the two conditions were a consequence of different error interpretations between the two conditions, and that different error interpretations occurred based on a the spacing cue.

Another cue to an error may be the consistency of the deviation between the automation's measurement and the actual measurement. Chronic, but consistently-off automation (e.g., always + 10% over-estimating) has higher trust and greater usage compared to automation that varies in error (Muir & Moray, 1996). In sum, consistency cues may tend to include how frequently an error happens and how much the error varies from the true value.

Context or distinctiveness-related cues to errors

People may also assign different interpretations to errors based on the simplicity of the situation in which the automation errs (Madhavan, Wiegmann, & Lacson, 2006). Madhavan et al. (2006) found that automation erring on simpler stimuli (containing more instances of the target and fewer non-targets) was trusted less than automation erring on complex stimuli (containing fewer instances of the target and more non-targets). To generalize this finding to a Fitbit One example, a user may think an error could occur on an unusually difficult task for the tracker or that the error may only occur in certain situations (e.g., running up a hill).

A separate experiment also found support for the cue of distinctiveness. In Masalonis (2003), some participants were told contexts (e.g., flight plan deviations) where the automation would not update, resulting in lower reliability on those trials. Those participants had a more appropriate relationship between trust and verification compared to participants not given context-specific information. This more appropriate relationship was characterized by increased verification when errors were more likely (when flight plans deviated) and decreased verification when the automation was less likely to err (when flight plans remained the same). Taken together, Madhavan et al. (2006) and Masalonis (2003) support the extension of the distinctiveness cue from person perception to automation error interpretation.

Explicit causal cues to errors

Few studies have provided causal cues to errors and none have assessed if participants found causal cues on their own. Instead, it has been assumed that participants used the provided cues based on the participants' different reactions between experimental conditions. For example, in Bagheri and Jamieson (2004) participants who were told that the technology would undergo updates every ten minutes were better at detecting errors than those not given such information.

When provided causal cues, people react to automation errors in unique ways, suggesting causal cues can indeed be utilized in error interpretation. In Bisantz and Seong (2001), one group of participants was told a decision aid might provide unreliable estimates due to enemy sabotage, a second group of participants was told the aid might err due to hardware and software problems, and a third group was not given any causal information. Participants in the sabotage condition reduced their use of the subsystem that could be sabotaged, but generally trusted the automation more than the hardware group. In another study, Dzindolet, Peterson, Pomranky, Pierce and Beck (2003) manipulated the explanation for the algorithm of a detection aid. They told some participants that the aid may occasionally mistake other shapes (e.g., the shading from a tree) as the target shape (e.g., a human), resulting in an error. Participants with this explanation trusted and relied on the detection aid more compared to participants without any causal information about automation errors. In sum, causal cues, and information provided explicitly to the users, may play a role in the interpretation of automation issues.

Logic-related cues to errors

Explicit causal cues may be particularly helpful because they provide the user with logic that supports a device mental model. Logic-related cues do not necessarily explicitly state the cause, but allow users to infer the cause on their own. Individuals with logic-related cues respond differently to errors depending on the extent to which an error reflects the intention of the automation's designer. Lees and Lee (2007) compared two qualitatively different types of automaton false alarm errors using driving automation simulations. One type of error was given the name “false alarm.” This type of error gave seemingly no reason as to why the alarm sounded (i.e., sounding at random). The other type of error, deemed ‘unnecessary alarms,’ also signaled a danger incorrectly, but with the reasoning for the incorrect alarm obvious (i.e., unnecessary but matching the user’s device mental model). In their study, an example of an unnecessary alarm would be an alarm occurring when a lead vehicle was decelerating to make a right-hand turn, but doing so at a rate where the participant’s car would just miss hitting the lead vehicle. Although unnecessary, this alarm could logically support a user’s device mental model because the situation the alarm sounded in was similar to situations the should sound in (e.g., situations where the participant would hit the lead vehicle).

Lees and Lee (2007) found unnecessary alarms were reacted to with more automation use compared to random false alarms. When an alarm correctly sounded to targets, those who had experienced false alarms responded with slower reaction times and lower brake frequency than those with unnecessary alarms. Additionally, those with unnecessary alarms had faster reaction times and higher brake frequency than those with 100% accurate alarms. The different reactions were assumed indicative of utilizing logic related cues.

Similarly, Rovira and colleagues provided information about the reliability for four specific components used by a decision-aid for each decision trial (Rovira, Cross, Leitch, & Bonaceto, 2014). Those users provided context-specific information performed better on automation failure trials than those not provided any such information. This further suggests that participants are able to understand the causes of errors (e.g., inaccurate information being used by the automation) or at least how information is fed forward in the automation, and adjust their reliance on the automation accordingly (e.g., cease use until the cause of the unreliable information was updated). In short, both explicit causal cues provided to the users and logical cues are cues that help develop device mental models and explain the causes of automation errors.

Other cues to errors

Thus far I have discussed the cues to errors found in the human-automation literature. However, other types of cues likely exist and may be particular to a given automation's functions. For example, most wearable activity trackers measure distance and sleep. Users may be able to use reference points as cues to errors for these functions. For example, users may run a 5K and compare it to their trackers. Or, a user may reference a clock to know he was awake for 15 minutes overnight and compare that information to his sleep tracker. Related to consensus, which has not been studied much in human-automation interaction literature, users might compare their tracker to other technologies. Or, users may simply have a feeling their tracker is wrong, such as with the tired user reviewing her sleep data. Clearly the cues to errors assumed in the human automation literature are not an exhaustive list of possible cues to errors.

To summarize, studies have manipulated explicitly providing participants with different types of knowledge or experimental conditions (cues to errors). However, there is an underlying assumption that this knowledge given to participants is akin to the sort of cues users would recognize from experience. It is unknown if users outside of experiments would utilize causal knowledge cues (and interpret the issue with a cause).

Error Interpretation

Cues to an error help users recognize an error is occurring and provide insight into what that error could be. In the studies discussed thus far, it is typically assumed the cue manipulation is responsible for the different response patterns of users, with error interpretation being the link between the cues to errors and the response. These assumptions tend be easy to find. For example, in Itoh et al. (1999), errors occurring in series may be interpreted as the result of an acute onset malfunction that can be fixed. In contrast, errors spread out over time might be interpreted as some sort of ongoing malfunction. However, these deep interpretations of errors did not necessarily have to occur. For instance, participants may have simply recognized an error was occurring, and without deciding on a cause, they may have simply decided not to rely on the automation when it was making frequent errors. As another example, in Dzindolet et al. (2003) the error was probably interpreted as being caused by the sensing algorithm as per what they were told. However, in all these studies that utilize cues to errors, participants did not explicitly state their interpretations. Therefore, there is no way to verify assumptions about what they thought was the error.

Furthermore, interpretations in everyday life are based on the cues to errors people find on their own. These cues may have merit, but they may not. The person perception literature suggests a perceived co-variation model wherein if the error occurs within the presence of a stimulus, but does not occur when the stimulus is absent, the stimulus must have caused the error (Kelley, 1967). However, this association could be an erroneous correlation (Chapman & Chapman, 1969; Golding & Rorer, 1972). Or, if no cues are available, people may not decide on any cause.

Even if cues are present, people may not even rely on causal cues and very detailed interpretation. For example, in the randomly occurring error studies (e.g., Riley, 1996), causal interpretation may never have occurred. Rather, participants may have simply recognized a general error occurred (e.g., something went wrong, but they did not know where it went wrong or what went wrong). For wearable activity trackers, people may interpret an error to a specific level, without a cause. An example of this would be a mistake or inaccuracy in the step count on the Fitbit One. Such an interpretation could be compared to a causal level of an algorithm that senses pendulum swings as a step. Indeed, it is also possible that in many of the situations where a person's device mental model and the situation's available cues to an error *could* allow the user to make a deep interpretation like a causal one, the user instead accepts a shallow level error interpretation (e.g., "something is wrong"). There are a variety of potential reasons for doing so, such as the error not having a substantial impact to the user, or the response being the same regardless of the error, or the amount of effort required to determine causality would be considerable. In short, is unknown to what extant individuals interpret automation errors in their everyday lives.

Issue Reaction

Reasons for Strategies

The human-automation interaction literature is lacking studies that examine how users decide to respond to an automation error. In most automation studies, few response options (discussed below) are available to participants in most automation studies. Furthermore, with so few response options available in studies, there is a lack of understanding as to why a user might respond in a particular why to an automation issue. Obviously, the lack of reasons for response strategies is especially inadequate for those possible strategies that have not explored in the literature. In other words, outside of the laboratory setting, where more reactions are feasible, more cues to strategies and more strategies likely exist.

Decades ago, before automation was as widespread as it is now, Konradt (1995) and Bereiter and Miller (1989) used interviews and observations to understand the response strategies of mechanical and electrical maintenance specialist troubleshooting failures in computer-controlled manufacturing systems. Although the computer-controlled manufacturing systems and activity trackers differ, and specialists and every day users differ, Konradt (1995) and Bereiter and Miller (1989) provided results that may transfer to everyday activity tracker users. These studies found that more than one reason for a strategy is often used in troubleshooting. Both studies also found ease of testing/least effort, historical information, and quality of the information obtained/information uncertainty as reasons specialists chose a particular troubleshooting strategy. In extending these cues to everyday activity tracker users, it seems likely that effort, experience with the activity tracker, the frequency of errors in the activity tracker's history, and the confidence in both that an error occurred and being able to determine what caused the error all may play a role in determining how to respond to an automation error.

Knowledge also may also help the user decide how to respond to an automation issue. It is feasible that cues to strategies may come from a user's device mental model, which includes relevant experiences not limited to the historical information of the activity tracker and includes the logical aspects of how the technology works. Previous experiences with other technologies or other errors may provide guidance in deciding how to respond. Additionally, knowledge from others may serve as a cue to a strategy. In interpreting an individual's behavior, people like to seek out information from other people to gain credibility for their own interpretations (Kelley 1967). Thus the cue "to see if anyone else had the problem" may emerge as a reason for looking up information on activity tracker error on online help forums. In sum, the reasons for selecting a reaction strategy to an automation error warrant further exploration as they likely influence the action taken by a user.

Strategies

Human-automation interaction studies have typically used measures related to use or verification as their results. Consequently, strategies participants can use to respond to the errors are limited to actions such as: cease using the automation for the experiment, cease using the automation temporarily, and monitor the automation. However, outside of the laboratory environment, many other strategies abound.

Indeed, if an automation error is considered a threat to the user's task of tracking his or her health, strategies used to manage threats to other tasks might emerge. For example, in managing health crises tasks, prevention, mitigation, work-around, and ignore strategies are common (Durso et al., 2015; Holden et al., 2012). In Konradt (1995) and Bereiter and Miller (1989) troubleshooting specialist always repaired the system (e.g., they would return the system to working condition after the error occurred). For activity tracker errors, there may be additional options such as investigating the error by trying to recreate it, replacing the tracker, or getting help from other users, online, or the company. Indeed, a taxonomy of automation error response strategies, and the cues that promote each strategy could guide training in error response. Wearable activity trackers provide a feasible means for developing such a taxonomy.

Activity Trackers

Over three million wearable activity trackers were sold between April 2013 and March 2014 alone (Danova, 2014). The popularity of activity trackers comes in the wake of two larger scale shifts of our everyday lives. One such shift is towards a patient-professional partnership in health care wherein individuals are becoming more educated and involved in making daily health self-management decisions (Bodenheimer et al., 2002). The second shift is one towards automation wherein automation has become ubiquitous. In light of these two shifts, it is not surprising that automation that supports health self-management is gaining in popularity (Bujonch, 2011).

Wearable activity trackers have a limited amount of automated features. These may include a step counter, stair counter, sleep tracker, calories burnt tracker, heart rate, clocks and alarms, and synching. A variety of automation imperfections can occur with any of these features and these trackers have mean absolute percent values caloric estimates that typically range from 10.1% to 13.0 % (Lee et al., 2014). When engaging in activities other than walking, the tracker's algorithms may inaccurately detect an activity as a different activity (mistaking biking for hiking), omit an activity (mistake biking for sitting), or provide credit for an activity not earned (e.g., mistaking riding in a car as running; Consolvo, McDonald, & Landay, 2009). These types of errors result in frustration for users (Consolvo et al., 2008). However, automation errors do not have to be limited to poor estimates. For example, if there is a hardware problem with the vibration only the silent alarm may not work, but the tracker could still be sensing steps walked. Or, hardware issue could occur to the entire device, such as water damage, resulting in failure to sense activity across a multitude of features. Some errors (e.g., an inaccurate algorithm vs. a complete hardware malfunction), may be more tolerable to users than other errors (Consolvo et al., 2008). Thus, not all error responses will be the same. In sum, activity trackers are an automation that can make a variety of errors and users could respond to these different errors in different ways.

Activity trackers have additional useful characteristics for investigating error interpretation and reaction. Although all automation errs, users of activity tracker may be particularly aware of these errors because users wear the tracker and can check its data frequently. Cues to errors may also occur because the automation is measuring the user's behavior. The users are, to some extent, aware of their own behaviors, and thus they have something to which to compare the automation's measurement.

Another reason to use activity trackers to examine error interpretation and reaction is that activity trackers are an example of an everyday technology that users learn about on their own. Outside from introductory welcome and quick-start guides, users have no formal training on the trackers, which is the case for many automated technologies (e.g., GPS). Research using case studies suggests that it takes between three days to months for users to develop a device mental model of one these technologies, and that comparing to self-behavior is one way users better understand the technology (Mackinlay, 2013). The experienced activity tracker user might draw on a mental model to troubleshoot automation errors in a similar way as for other everyday automation learned without formal instruction. In sum, activity trackers are a useful medium to examine automation error interpretation and reaction.

A Theoretical Model for Automation Issue Troubleshooting

Based on the literature of human-automation interaction, device mental models, person perception, and activity trackers, I proposed a theoretical model of how people interpret and react to automation issues (Figure 1). The proposed framework identifies gaps in the literature surrounding error interpretation and reaction.

Issue Interpretation

The first step in the proposed model is attending to a cue to an automation issue, and interpreting it as indicative of issue. Additionally, the user must perceive this issue to actually be an error, and not part of the intentional design, if the user wants to try to remedy the situation. An example of a cue to an error might be if the step count has not changed over the course of several hours. Through a user’s device mental model, the framework takes into account the role experience and knowledge might have in serving as additional cues to errors. In other words, the first cue perceived (e.g., a discrepancy between an expected number of steps walked and the

Hypothesized Framework For Automation Issue Troubleshooting

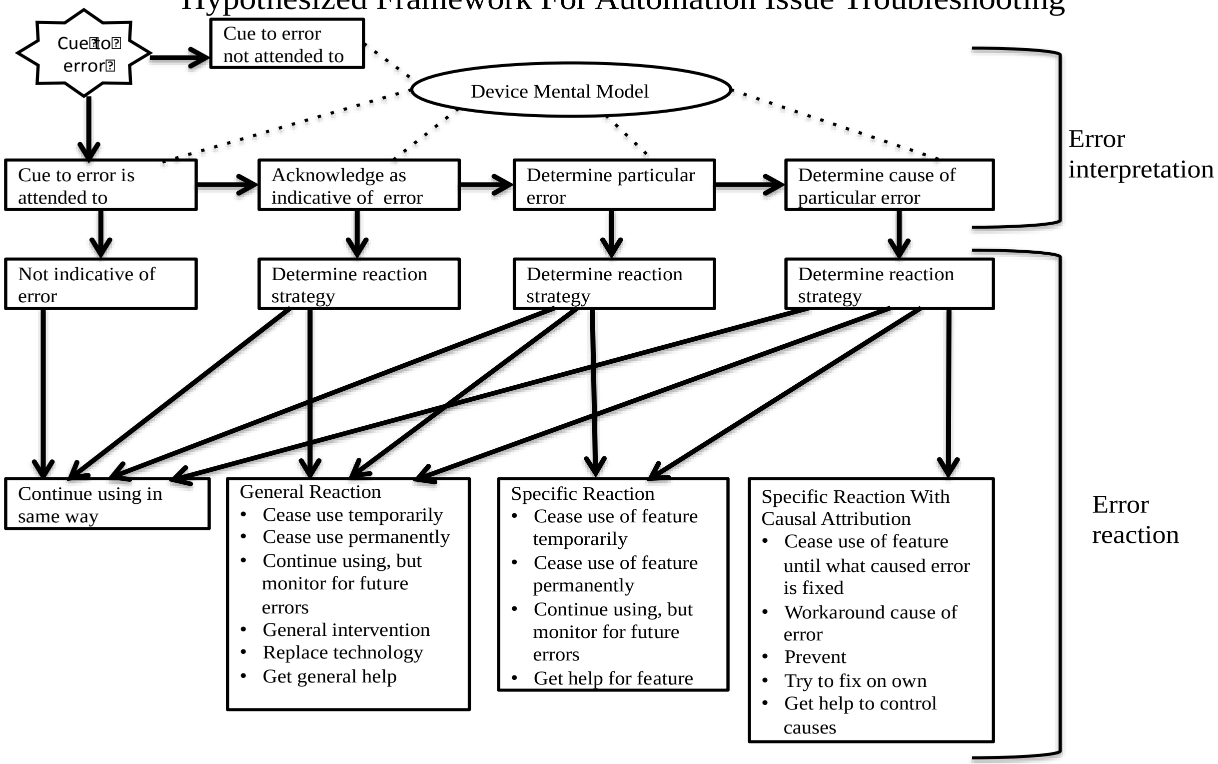


Figure 1. Proposed initial model of how individuals interpret and react to automation issues. Attending to a cue to an issue gives users the opportunity to classify the cue as pertaining to an error or not. If the user acknowledges the cue as indicative of an error, the user may not determine what specific error occurred (e.g., "something is wrong") whereas particular errors can also be determined (e.g., "the step counter is not registering movement"). If users choose to interpret the error further by hypothesizing its causes, they may do so (e.g., "the step counter is not registering movement because the sensitivity setting of the tracker is not sensitive enough.") Cues that promote the development of a user's device mental model (e.g., consistency, consensus, distinctiveness) may help in error interpretation. For example, distinctiveness may help users determine causes (e.g., "if the device only errors when charging, the device might be shorting out because of loose connections"). The strategies available to react to an automation error may depend on the extent the user interprets an error. General reactions are broad strategies that only require acknowledging an error occurred (e.g., a general intervention may be restarting the technology). Specific reactions require specifying the error. Hypothesizing causal attribution allows users to react at the level of the cause (e.g., "change the sensitivity settings so that the step counter will register movement.")

automation's reading of the number of steps walked) is not necessarily the only cue a user may have in interpreting errors. For instance, participants might draw on experiences over situations, over time, and over different technologies if a perceived error has cues of distinctiveness, consistency, and consensus. To elaborate, imagine that a user's Fitbit One rarely fails to count steps accurately with the exception of when the tracker has been outside in the sun for several hours. However, when it has been outside in the sun for several hours, the activity tracker almost always does not count any steps taken. Further, imagine that the user has utilized multiple wellness management technologies and knows that the Nike Fuel Band and the Striiv make similar errors. Given all these cues in addition to the discrepant step count, the user may hypothesize that the cause of the error is the technology over-heating.

Drawing on the research in which humans differentiate their responses to automation errors based on the cues to errors, a possible next step in error interpretation is to differentiate or not differentiate that error. For example, the user may simply acknowledge that there is an error without knowing what is impacted by that error (e.g., "something is wrong with my activity tracker"). However, the user might interpret the error more specifically. In other words, the user understands not just that *an* error has occurred, but *where* that error occurred or *what* is that error. For example, users of the Fitbit One or the Jawbone Up 24 may perceive no changes in their activity log as "something is wrong" at the "acknowledge as indicative of error" detail category, but in the "determine particular error" category, the user might become more particular and instead think "the step counter is not registering movement." Although in automation studies, people have been given causal information, it is unclear if people hypothesize and intervene at the causal level on their own when resolving automation errors. However, given that people could hypothesize causes, an even more detailed interpretation than knowing *where* the error occurred would be knowing *why* the error occurred. In continuing with the previous example, in the causal level of interpretation of "determine cause of particular error" a person may reason that "the step counter is not registering movement because the sensitivity setting of the tracker is not sensitive enough." Presumably, some cues to errors (e.g., experience) might help in interpreting error in more detailed.

Issue Reaction

The proposed model assumes that the extent to which an attended to error is interpreted will determine the repertoire of strategies an individual has available to respond to that error. If an error has only been interpreted generally, than only broad tactics are available to cope with the error. One example of this would be to restart the technology. If an error is interpreted to a more specific level, additional strategies become available. For example, perhaps the user has pinpointed the error as only being on the sleep-tracking feature. Then the person may target just that feature in reaction by continuing to use the step counter, but ceasing use of the sleep-tracker. Although some cues to strategies such as ease of implementation have been found in the trouble-shooting literature (e.g., Konradt, 1995), there are no known or hypothesized links as to with which error interpretation details or with which strategies certain cues to strategies may emerge.

Finally, if the user has attributed the perceived error to a cause, the user has the option to intervene at the causal level as well. One strategy of this category may be trying to prevent the error from occurring. In returning to an earlier example, if the user believes errors occur from the technology over-heating, the user might try to leave the technology in the shade instead of the sun. These strategies might be similar to other common strategies for handling threats to a task (e.g., Durso et al., 2014). Earlier strategies also could remain available at the later stages of error interpretation. For example, even if the user did think the errors occurred from over-heating, the user may still choose just to restart the technology instead.

Initial Limitations of Theoretical Framework

The proposed framework provided a starting point for understanding the automation issue troubleshooting process. However, because it was a starting point, there were several limitations that should be considered. First, although they are arranged in series, there is no reason why a user must first interpret an issue generally to be able to interpret in causally. Additionally, the possible categories of error interpretations and possible reaction strategies are not limited to those discussed in the proposed framework. The purposes of this framework was to summarize the findings and gaps in the literature, guide research questions, and guide analyses of automation error interpretation and reaction. Consequently, some of the proposed framework may need to be revised. Findings from this thesis study help revise the theoretical framework.

Based on the gaps in the current literature, the particular research questions for this study were:

- [R1] What information serves as a cue to an automation issue for experienced everyday automation users?
- [R2] With what level of detail do experienced everyday automation users interpret automation issues?
- [R3] Upon attending to an automation issue, how do experienced everyday automation users decide to respond?
- [R4] What strategies do experienced everyday automation users have for responding to an automation issue?

- [RS] To what extent do users' device mental models relate to how they interpret automation issues?

Overview of Study

The present study took a multi-method approach to qualitatively examine how people interpret and react to attended to automation issues. In three interviews, experienced activity tracker users discussed troubleshooting their activity trackers. The first interview built rapport. The second interview combined a threat-strategy interview (TSI; Durso Kazi, & Ferguson, 2014) with a critical incident interview (Flanagan, 1954). This second interview allowed cues to errors, error details, cues to strategies, and strategies to emerge. Methods of this nature have proven reliable (e.g., Andersson & Nilsson, 1964) and ideal for exploratory research (Woolsey, 1986). The third interview was a scenario based interview and it ensured some level of comparability across participants. The scenario based interview kept the cue-to-errors constant across participants, and allowed for error interpretation details, cues to strategies and strategies to emerge. Multiple questionnaires were also utilized to obtain pertinent background information (e.g., device mental model, technology experience, automation-induced complacency).

CHAPTER 2

METHOD

Participants

Thirty participants (12 males) were recruited from courses, advertisements in campus newspapers, campus flyers, email lists, and word of mouth. Participants were compensated with credit towards a course or $20. To be eligible, participants were required to have used their own wearable activity tracker at least three days a week for the four weeks prior to their participation. Participants were required to have access to their respective device and its phone app or website while participating in this study, with the exception of during device mental model assessments. Participants ranged in age from 18 to 56 years (M_{age}=25.43, SD_{age} = 7.45). Nineteen participants described their race as Caucasian, six as Asian, three as Multi-racial, and two as African-American. Fifteen participants reported a formal education of a bachelor's degree or above, and 15 participants were undergraduate students.

Materials

Questionnaires were administered to describe the sample population and to assess relevant knowledge and experience. Interviews were administered primarily to elicit cues-to-errors, errors, cues-to-strategies, and strategies. Table 1 summarizes the assessments used in the final version of this study.

Wellness Management Technology Background Questionnaire

This questionnaire was designed for this study to assess participants' uses of, and motivation for using, wellness management technologies (Appendix A). Through multiple choice, open ended, and Likert-type questions, participations reported the other activity tracking

Table 1
Assessments Used In The Current Study

Data Assessed	Measures
Cue to errors, errors, strategies, and cues-to-strategies	Threat-Strategy Interview/Critical Incident Interview
Errors, cues-to-strategies, and strategies across the same cues-to-error across all participants	Scenario-based Interview
Expertise of activity tracking technology	Wellness Management Technology Background Questionnaire; Device Mental Model Knowledge Questionnaire; Activity Tracker Explanation Form
Use of activity trackers and motivation for use	Wellness Management Background Questionnaire; Technology Experience Profile 48-63/64
Perceptions on automation	Wellness Management Technology Background Questionnaire; Automation Measurement Profile
Technology experience	Technology Experience Profile questions 1-36
Demographics and background	Background and Health Questionnaire
Exercise habits	CHAMPS Questionnaire

technologies they had used and, based on Fox and Duggan (2013), reported how they kept track of their activity before using a tracker. Participants also explained what motivated them to use their activity tracker, and were asked to rate their motivation to use their trackers for fun and for health reasons. Participants were also asked how long and how frequently they used their activity tracker, including how often they checked it information. Lastly, participants also reported their general perceptions about the accuracy of their activity trackers.

Interview Definitions

During the three interviews, participants referenced definitions that represent automation, automation errors, and strategies for responding to automation errors. The definition for automation as adapted from Moray et al. (2000, p.44) explanation of "any sensing, detecting, information-processing, decision making, or control action that could be performed by a human but is actually performed by a machine." Automation errors were described broadly as issues that could make it difficult for users to track their activity, fitness, or health accurately. The definition was loosely based on the Consolvo et al. (2008) documentation of a different type of activity tracker wherein participants described activities the tracker inaccurately inferred. This definition also allowed participants leeway in determining if a device limitation or human error was an automation issue, but not necessarily an automation error. Relevant examples were given for both "automation" and "automation issue" in their definitions. "Strategy" was also defined in accordance with Durso et al. (2014, p.4) and the description explained to participants that a strategy is a means to complete a goal and may sometimes be thought of as an action.

As a caveat, because materials development and pilot testing revealed very strong connotations around the words "automation" and "automation error" we avoided those words in the definitions given to participants. Instead, automation was labeled "Activity Tracker's Function of Sensing Detecting and Information Processing" and "automation error" was expanded and labeled "Activity Tracker Issue" for participants. Table 2 lists the complete definitions given to participants and what constructs those definitions represent. Participants were asked to focus on these definitions, such as issues of features that met the automation definition, during the interviews.

Introductory Interview

The Introductory Interview was designed to encourage participants to start thinking about the automation errors made by their activity trackers and consisted of general questions related to

Table 2
Definitions Used in the Current Study

Construct	Label Given to Participants	Definition Given to Participant	Relevant citations
Automation	Activity Tracker Functions: Sensing, Detecting, and Information Processing	In this interview, we will talk a lot about what your activity tracker does automatically for you, on its own. Think of all the things your activity tracker does for you, that you do not have to do, such as sensing, detecting, or information-processing. • For example, with sensing or detecting, if you wanted to, you could count every step you take throughout the day. Or, technology like your activity tracker, can count every step you take throughout the day. • Likewise, as an example of information processing, you could calculate how many calories you burned today, or your activity tracker can do it for you. Although your activity tracker has many features, we are going to focus on these components where you could do what your activity tracker does, but it does it for you.	Moray et al. (2000)
Automation Issue (which includes all Automation Errors)	Activity Tracker Issues	Think about times when your activity tracker was supposed to do something for you, but you have think it did something wrong. • For example, maybe the activity tracker provided information that differed from what you thought it should say under those circumstances. • Or, your activity tracker may have made a mistake with sensing, detecting, information processing, or any thing the activity tracker does on its own. This might include situations where it made a very particular error on a feature or you thought something was going on, even if you did not know exactly what is was.	Consolvo et al. (2008)
Strategy	Strategy	A strategy is a plan or method to achieve a goal. A strategy is not usually one action, but you may think of it as an action.	Durso et al. (2013)

activity tracker usage, understanding, and perceptions. Additionally, the definition of automation was also introduced to participants during this interview. Because the purposes of the Introductory Interview were not focused on the actual interpretation and reaction to automation errors, its data are not reported here.

Threat-Strategy/Critical Incident Interview

The primary purpose of the second interview was to capture salient potential automation issues to which users interpret and respond. This part of the interview corresponded to the 3 stages of the threat-strategy interview (TSI) methodology (Durso et al., 2014). TSI questions were integrated with questions from the critical-incident (CI) methodology (Flanagan, 1954; McBride, 2014), particularly in the second stage. The combined TSI/CI Interview (hereafter referred to just as TSI) allowed for cues-to-errors, errors, cues-to-strategies, and strategies to emerge.

<u>Stage 1</u>

Participants first received an introduction to the critical task. The critical task was broadly defined to participants as using an activity tracker to keep track of their activity, fitness, or health. Participants were asked to explain how they would perform the task. The primary purpose was to provide a goal that could be hindered by automation errors. Because the purpose of this stage was to provide context, it was not coded and that data are not reported here.

<u>Stage 2</u>

The second stage elicited threats (i.e., errors) and cues-to-threats (i.e., cue-to-errors). Participants were introduced to the definition of automation issue. Then, participants generated a list of issues and each issue was later discussed in greater detail one-by-one. For each automation issue, participants were asked to describe the nature of the issue. This question typically allowed for the details (e.g., general, specific, causal) of the issue interpretation to be captured. Next, participants were asked how they became aware of the issue and how the cues they mentioned were related to the issue. Participants were then asked when the issue occurred. Finally, participants were asked about what occurred prior to the issue.

Stage 3

The third stage of the TSI extracted strategies and cues-to-strategies. At the start of Stage 3, participants were given the definition of strategy. Then, for each issue, participants were asked for a strategy to keep that issue from interfering with tracking their activity, fitness, or health. Participants justified their choice of strategy. If participants provided an intervening strategy, they were asked how they would know if that strategy worked. Participants were then asked for another strategy, and/or when relevant, what they would do if their first strategy did not work. Participants were asked to justify why they would try this next strategy. Elicitation of strategies continued until the stopping rule. The stopping rule was defined as “discontinue use permanently, not to track, to track wrong, to send back to manufacture, to manually override the automation, or to track wrong and correct in your mind.”

Typically, participants described a situation that had actually happened to them. If the situation had happened to them, additional questions were asked about what they actually did in the scenario, which sometimes resulted in the revelation of additional strategies. Participants were then asked to weigh all the strategies they could have or did use against one another, which helped to reveal additional cues-to-strategies. Lastly, participants were asked if any previous experience was helpful to them in thinking about how they could deal with the situation. This question also helped to reveal cues-to-strategies. For each issue generated by participants, Stage

2 and Stage 3 questions were asked in series prior to starting Stage 2 for the next issue. The entire TSI is in Appendix B.

Scenario–Based Interview

The Scenario Based Interview (SBI) was to designed compare issue interpretation and reaction across participants. Cues-to-errors were held constant in five scenarios that were given to participants. This allowed for issue details, cues-to-strategies, and strategies to emerge. The questions of the SBI were similar in nature to those of the TSI. However, the SBI had fewer questions related to cues-to-issues, as cues-to-issues were given in the scenario prompt, and did not have any questions related to what the participant actually did, as the scenarios were hypothetical. The SBI also differed from the TSI in that the SBI had two quantitative questions: one related to the impact of the issue and one to the confidence in classifying the scenario as an issue or not an issue.

Each of the five scenarios was developed for this study. Some hypothesized cues-to issues were incorporated into the scenarios. For example, distinctiveness cues occurred in naming the specific activity of an indoor spin class in Scenario A. Although not asked explicitly, the cues could then encourage the participant consider if the automation error occurred during other activities (i.e., activities when their arms and activity trackers may be still). Continuing with this example, consensus was also incorporated as an indoor spin class necessitates a stationary bicycle machine and exercise machines typically provide measurements related to the activity, such as calories burned. Comparing the activity tracker to the machine would be an example of consensus.

These five scenarios were designed to capture situations that could happen to almost every popular wearable activity tracker. Therefore, automatic features common across activity trackers initially limited the options of scenarios (e.g., calorie tracking, distance tracking, step tracking, sleep tracking, and the wearable tracking device as a whole). Online FAQ and help forum searches and pilot testing were used to develop realistic scenarios. These scenarios also came up in the TSI portion of the interview, suggesting that they were indeed valid scenarios.

The scenarios were created to be applicable to similar error situations (e.g., inaccurate sensing during a spin class is similar to inaccurate sensing during a car ride). Nonetheless, multiple, very different, errors could be plausible from the same scenario. This variability was sometimes due to strong device mental models coupled with unique features of specific activity trackers and sometimes due to the intentional vagueness of the scenario. The expected possible error interpretations captured for each scenario are described in Table 3.

At the start of the SBI, participants were reminded of the definitions of "Activity Tracker's Function of Sensing Detecting and Information Processing" (automation), "Activity Tracker Issues" (which includes automation error), and "Strategy," and were introduced to the two quantitative questions and corresponding scales. The first scale was used to judge the impact of each scenario on the participant's ability to tracking his or her activity, fitness, or health using his or her activity tracker. Activity, fitness, and health were all mentioned because individuals' goals for using an activity tracker might differ. Although some impact scales have been recommended in human factors' usability (Sauro, 2013) no clear scale is preferred and many are not appropriate for the activity tracking task at hand. Thus, a scale used to measure the impact of problems across many fields was chosen (e.g., behaviors at school; Williams, 1992; illness symptoms; Debeau & Versi, 2001). This scale has 4 anchored points: no impact, minor impact, moderate impact, and serious impact. The scenarios ranged in impact, with most scenarios being

Table 3

Design of the Scenarios in the Scenarios Based Interview

Scenario	Automation Issues Captured
A. Imagine you have just completed a rigorous indoor spin class. However, your activity tracker has not substantially increased the number of calories you have burned.	• Scenarios with sensing and information processing design imperfections for activities (i.e., no credit when tracker is still) • Incorrect basal metabolic rate calculations • Inaccurate sensing (e.g., wrong sensitivity setting, software update) • Synching problems
B. Imagine you have just walked all of Georgia Tech's Pi Mile trail, which you know to be exactly 3.14 miles. You notice that your activity tracker says you have walked 3.5 miles.	• Distance calculation discrepancy or inaccuracy based on number of steps (e.g., difference between mapmyrun and fitbit) • Inaccurate or variable stride-length, based on height • Incorrect count of stair cases climbed is a similar scenario • Wearing on the wrong wrist/sensitivity settings in which steps are over counted • GPS failure to integrate (varies by activity tracker)
C. Now imagine the very same situation. You have just walked all of Georgia Tech's Pi Mile trail, which you know to be exactly 3.14 miles. You notice that your activity tracker says you have walked 3.5 miles. However, you know this occurs every time you walk the Pi Mile trail.	• Same as Scenario B, but consistency suggests ongoing problem (e.g., wrong stride length used in calculation)
D. Imagine you are reviewing your sleep data and find that your activity tracker says you only awoke once throughout the night. However, you recall waking up several times.	• Inaccurate information processing of minimal, but sensed movement (e.g., if there is not a prolonged instance of movement, an app deems being awake as restless sleep; heart rate changes to sense peak activity in exercise is a similar situation) • Incorrect sensing of movement • Potential synching problem
E. Imagine you can not get the wearable tracking device to respond at all, even after charging it. When you press the button on the device itself, none of the displays (including any screens or lights) appear. So, the device is not responding at all.	• Complete malfunction (e.g., water damage, smashed, heat damage) • Manufacturer's defect, hardware problem • Lights or button broken • Cannot charge/hold a charge

designed to have a moderate, but not absolute, impact on the ability of a user to use his or her activity tracker. However, one scenario was developed to have a severe impact (Scenario E).

The second scale was developed for this study. The purpose of the scale was twofold: (1) determine confidence in how the participant had assessed the scenario and (2) facilitate the interviewer selection of appropriate follow-up questions. An even-numbered bi-polar scale was designed with 6 anchored points: extremely confident the scenario is not an issue, moderately confident it is not an issue, slightly confident it is not an issue, slightly confident it is an issue, moderately confident it is an issue, and extremely confident it is an issue. Both the impact and confidence scales are in Appendix C.

After reminders of critical definitions and introductions to the scales, participants were given each of the five scenarios one at a time.

The five scenarios were presented to participants in order of their concreteness. This was because most participants had only discussed issues that had actually happened to them, as opposed to hypothetical issues, in the TSI/CI Interview. For each scenario, participants received a piece of paper with the scenario written out and the interviewer read this scenario. Then, participants were asked to rank the impact of this scenario on their ability to track their activity, fitness, or health on the 4-point Likert-type impact scale. Participants were asked what they thought was going on in the scenario and if they thought it was an issue. Participants ranked how confident they were in their judgment of the scenario as an issue or not an issue on the 6-point Likert-type confidence scale. If the confidence leaned more towards being an issue, questions were asked about what the issue was, and if applicable, what suggested a cause or specification was the root of the issue. If participants leaned in their confidence towards the scenario not being an issue, they were asked why they did not think the scenario was an issue. These questions typically revealed the details of how the error was interpreted.

All participants were asked how they would respond to the situation and why. If applicable, participants were also asked what they would do if their strategy did not fix the problem. These questions revealed strategies and cues-to-strategies. The entire SBI is in Appendix D.

Activity Tracker Explanation Form

This questionnaire (Appendix E) was designed for this study as a method to elicit the elaborate details of a user's device mental model of an activity tracker. Respondents were told the purpose of the questionnaire, as recommended by Van der Veer & Melguizo (2003). Next, respondents were asked to list all the features of their activity tracker. Then, because both written and diagramed explanations have been used to assess device mental models (Van der Veer & Melguizo, 2003), both methods were employed on this measure. Users were asked to explain how their activity tracker works and to draw how the entire activity tracking system (i.e., wearable device, app, website) functions. In addition to the scoring guide explained below, keys were created for each activity tracker used in the study (e.g., a Fitbit Flex key of all possible features, a Fitbit Charge key, etc.). Scoring keys were created from personal experience and from researching each activity tracker online. These keys gave scorers an idea about what to expect when applying the coding scheme and could be consulted to help determine if any details participants gave were inaccurate.

Two scores were generated for each participant's activity tracker explanation form. The first score was designed to capture the elaborateness of a user's device mental model. One point was awarded for each feature or function participants listed. Additional points were added for key words (e.g., GPS) in explanations of how the activity tracker works. The coding scheme for these key words is in Appendix F. Additional points were added for each link made in the diagram of the activity tracking system. An example of a link would be an arrow going from the wearable tracking device to a cellphone via Bluetooth. Thus, higher scores represent more elaborate device mental models.

The second score was designed to capture the inaccuracy of the user's device mental model. For this score, one point was added for any non-redundant information provided that was inaccurate. An example of an inaccurate fact would be describing heart rate as a feature when the activity tracker does not have a heart rate measure. Thus, lower scores represent more accurate device mental models. However, only four participants stated one wrong fact. In other words, participants typically did not include inaccurate information in their device explanations and inaccuracy scores ranged only from 0 to 1. Therefore, we did not further examine this score and it is not reported in the results.

Device Mental Model Knowledge Questionnaire

This questionnaire was designed for this study to assess the accuracy of participants' device mental models. Participants first identified their wellness management and then they answered ten True/False statements about the automated features of their activity trackers (Appendix G). An example of an item is "My technology uses an altimeter." Different keys were created for each activity tracker used in the study because some of the ten items were true for one tracker (e.g., Fitbit One) but false for a different tracker (e.g., Jawbone Up). Keys were created from researching each activity tracker online. One point was awarded for each correct answer such that higher scores represent a more accurate device mental model.

Technology Experience Profile

The Technology Experience Profile questionnaire was designed to assess familiarity with 36 different technologies (e.g., mobile phone, automated teller machine) on a scale from 1 (not sure what it is) to 5 (used frequently; Barg-Walkow, Mitzner, & Rogers, 2014; Appendix H). When scored, the Technology Experience Profile has general technology breadth score ranging from 0 to 36 where higher numbers represent a greater range of different technologies a person uses.

This measure was adapted to include additional automation (Johnson, 2004), wellness management technologies, and the different features of activity trackers. For example, we queried participants the items “the phone app” and “the alarm” of their current activity tracker. The same 5 point scale and format as the Technology Experience Profile were used for these additional questions. However, these additional items are reported on their own and are not included in the Technology Experience Profile breadth score and frequency profile.

Automation-Induced Complacency Potential Questionnaire

The Automation-Induced Complacency Potential Questionnaire, originally developed by Singh, Molloy, and Parasuraman (1993), and updated by Pop and Stearman (2015), assessed users’ biases toward automation usage (Appendix I). For simplicity, this questionnaire was called the Automation Experience Questionnaire when it was given to participants. In particular, it measures the amount of suspicion an individual has towards automation. When summed, scores can range from 16 to 80, and higher scores represent less suspicion of automaton. Sixteen automation questions are asked on a 1 (strongly disagree) to 5 (strongly agree) Likert-type scale. An example of an item is “Manually sorting through emails is more reliable than computer-aided searches for finding emails in my inbox.”

CHAMPS Questionnaire

Eighteen questions from the Community Healthy Activity Model Programs for Seniors questionnaire of Stewart et al. (2001) were used or adapted to assess the physical activities participants engaged in on a weekly basis, and how much time participants spend doing those activities (Appendix J). Three additional questions relevant to activity trackers were also added to the modified CHAMPS questionnaire and included the use of elliptical machines, the use of stair machines, and weekly exercise classes. In adapting the questionnaire, activities were added or removed if those activities were related to known situations in which activity trackers may err or if those activities were related situations that may provide additional insights into the workings of activity trackers. Twenty specific items of activities were included. However, participants could add activities not otherwise captured at the end of the questionnaire. The questionnaire required up to three answers by participants: (1) if they did or did not partake in that activity in the last four weeks on a weekly basis, (2) if they did partake, how many times per week they did so, in free response format, and (3) if they did partake, how many hours they did so, in multiple choice format. Scores reported in the results reflect the number of activities in which participants partook during a typical week.

Background and Health Questionnaire

The Background and Health Questionnaire was adapted from Czaja et al. (2006) to gather demographic information and self-reported health information (Appendix K). Most of the questions were multiple-choice, with some free response exceptions (e.g., date-of-birth, height, college major). Four health-related questions utilized 5-point Likert-type scales, with the anchors varying per question. Because the way an individual views and uses a technology may be dependent on their personal needs, dieting and prescribed exercises questions were included.

Automation Measurement Profile

This questionnaire was designed for this study to (1) determine users' perception of their activity trackers' accuracy and (2) to determine how consistent this perception is across various features (Appendix L). Participants were given a reading from their activity tracker. An example reading is "your activity tracker reports you have walked 7,000 steps." Participants checked off as many boxes they believed could represent the true amount (e.g., how many steps they have actually taken). Because earlier wearable activity trackers have been shown to range from 9.3% to 23.5% in their calculations of calories (Lee et al., 2014), 10% was used as a midpoint amount to convert error readings by, with 5% and 15% also included. Thus, between 1 and 7 answers were possible for each question: underestimation by 15%, underestimation by 10%, underestimation by 5 %, complete accuracy, overestimation by 5 %, overestimation by 10%, and overestimation by 15%. Returning to the example item, and example response might be a participant checking 4 of the 7 possible answers: 6,650 steps (5% underestimation), 7000 steps (complete accuracy), 7,350 steps (5% overestimation) and 7,700 steps (10% overestimation). This would suggest the participant thought that for the steps walked, the tracker was not completely accurate but not extremely inaccurate (e.g., 15%), and was more likely to have larger overestimation than underestimation.

Procedure

Participants were pre-screened via email to ensure they qualified for the study. Participants brought their wearable activity tracker to the Human Factors and Aging Laboratory, where they swere given an overview of the study, provided informed consent, and showed their activity tracker to the interviewer. Next, participants completed the Wellness Management Technology Background Questionnaire. The three interviews followed and were audio recorded for transcription purposes. The interviews occurred in the following order: (1) Introductory

Interview, (2) TSI/CI Interview, and (3) SBI Interview. The remaining questionnaires were administered after the interviews in this order: (1) Activity Tracker Explanation Form, (2) Device Mental Model Questionnaire, (3) Technology Experience Profile, (4) Automation-Induced Complacency Potential Questionnaire, (5) CHAMPS Questionnaire, (6) Background and Health Questionnaire, and (7) Automation Measurement Profile. Finally, participants were debriefed and compensated for their time. Figure 2 diagrams this procedure and the experiment script is available in Appendix M.

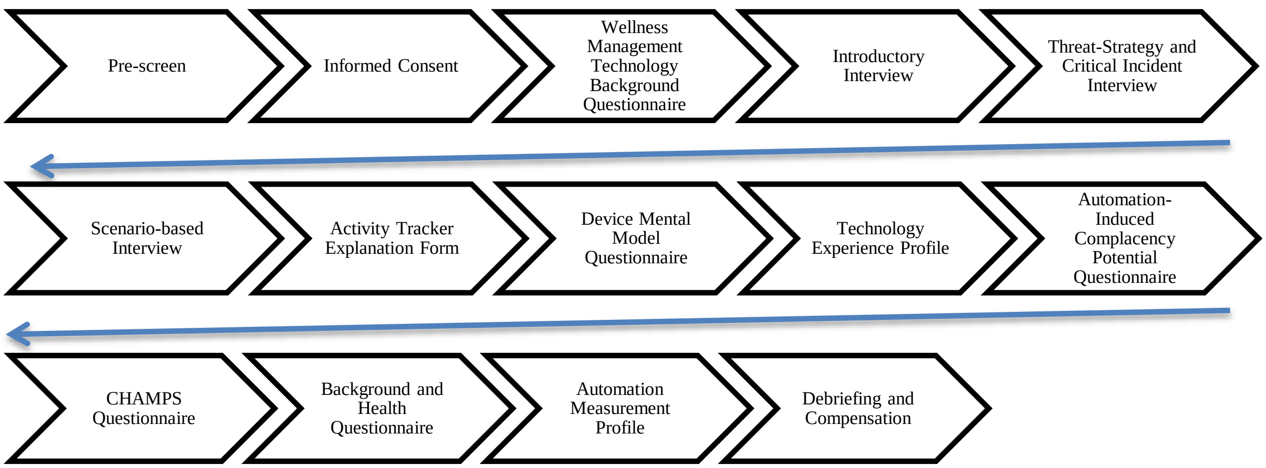


Figure 2. Procedure flow.

Design

Participants brought their own wearable activity tracker to the study. These trackers were 10 Fitbit Charge HRs; 9 Fitbit Flexs; 4 Fitbit Charges; 2 Fitbit Ones; 1 Jawbone Up 24; 1 Microsoft Band; 1 Garmin Vivofit; 1 Motorola 360; and 1 Fitbit Surge. Natural-groups variables related to participants' experiences and perceptions, in particular the accuracy and elaborateness of their device mental models. Dependent variables included: (1) the cues-to-issues that participants generated in the TSI/CI Interview, (2) the details of the explanation of the nature of the errors attended to (e.g., to causal) in the TSI/CI and SBI interviews, (3) the cues-to-strategies that participants provided in the TSI/CI and SBI interviews, (4) the strategies that participants provided in the TSI/CI and SBI interviews, (5) the impact of each scenario on participants in the SBI, and (6) the participants' ratings from the impact and confidence scales in the SBI.

CHAPTER 3

DESCRIPTIVE RESULTS

Subjective Health and Health Activities

We assessed participants' subjective health and activities to provide background and context for their experiences and knowledge of their activity trackers. As a whole, the sample was reportedly healthy and active. On average, participants reported their health as "good" or "very good" (Table 4). When asked how satisfied they were with their health, participants averaged between "neither satisfied nor dissatisfied" and "somewhat satisfied." Table 4 includes histograms for these two questions, where responses of 1 (poor/not at all satisfied) to 5 (excellent/extremely satisfied) are represented on the horizontal axis and frequency is the vertical axis. Thus, although the full range of the scale was used, answers clustered between 3 and 5. Twenty-one participants were not on any diet. Of the nine participants on diets, six participants had diets only related to general healthy eating habits (i.e., lean meats, watching calories, eating many vegetables) and 3 participants specified gluten-free or dairy-free diets. Only one

Table 4
Descriptive Statistics of the Health of Participants

Descriptor	**M**	**SD**	**Range**
Self Reported Health [a]	3.97	.91	1-5
Health Satisfaction [b]	3.90	1.08	1-5
Number Weekly Physical Activities [c]	8.27	2.75	3-14

Notes:
[a] On a scale of 1 (poor) to 5 (excellent), reported on the Background and Health Questionnaire
[b] On a scale of 1 (not at all satisfied) to 5 (extremely satisfied), reported on the Background and Health Questionnaire
[c] Total number of weekly physical activities reported by participants on the modified CHAMPS Questionnaire

participant described doing a prescribed exercise (physical therapy for a dislocated knee).

As a whole, participants were fairly active. On average, participants took a part in about eight different physical activities on a weekly basis. The most frequently reported activities were to “do work around the house (such as washing windows, sweeping, vacuuming)” (n=27), “walk leisurely to do errands, to attend classes or meetings, to exercise, or for pleasure” (n=27), “walk uphill or hike uphill (count only uphill part) outside” (n=25), “do strength-training exercises (such as hand-held weights, weight machines, or push-ups)” (n=19), and “go for a jog, run, or sprint session outside or on a track” (n=18). As a whole, our sample was fairly active.

Technology Experience

General and Automation Technology

Participants had experience with a wide range of technology, as indicated by their Technology Experience Profile General Technology Breath score in Table 5. We also assessed participants’ suspicions of automation. As a whole, participants tended to be somewhat unsuspicious of automation, as indicated by their automaton-induced complacency score (Table 5).

Table 5
Descriptive Statistics of Participants’ Technology Experience

Descriptor	M	SD	Range
General Technology Breadth [a]	28.73	2.85	23-34
Automation-Induced Complacency [b]	60.53	7.99	42-74

Notes:
[a] Potential scores range from 0 to 36, with higher numbers representing a greater variety of technology experience. Reported on the Technology Experience Profile.
[b] Scores range for 16 to 80, where higher numbers represent less suspicion of automation, as reported on the Automation-Induced Complacency.

Wellness Management Through Activity Trackers

Prior Wellness Management Experience

Fourteen participants (47%) reported that their current activity tracker was their first wellness management technology. Of the 16 participants (53%) with previous experience, both fitness applications and other wearable trackers were used. The types of apps used included myfitnesspal, gymboss, fullfitness, Nike +, Nike+ shoes, mapmyfitness, fitstar, mapmyrun, the iphone health app, and the Samsung s health app. Other wearable trackers used included the Jawbone Up, Fitbit Zip, Fitbit Charge, Fitbit One, and pedometers. Prior to using a wellness management technology, 17 participants (57%) were not keeping track of their fitness. Of the 13 participants (43%) keeping track prior to using a wellness management technology, 10 (30%) were doing so in their heads, two used paper, and one used a spreadsheet.

Motivation for Using an Activity Tracker

Free response questions varied in specificity and typically mentioned why a participant chose his or her particular tracker compared to other trackers. However two Likert-type questions from the Wellness Management Technology Background Questionnaire shed light on the extent to which the participants used their trackers for fun and the extent to which they used their trackers for health and wellness. Participants used their activity trackers about as much fun as for health (Table 6).

Table 6

Descriptive Statistics of Participants' Motivations for Using an Activity Tracker

Descriptor	**M**	**SD**	**Range**
For Fun [a]	4.63	1.49	1-7
For Health and Wellness [b]	5.27	1.36	2-7

Notes:

[a] Scale ranges from 1 (not at all for fun) to 7 (completely for fun), with higher numbers representing a greater amount of motivation for fun. Reported on the Wellness Management Technology Background Questionnaire.

[b] Scale ranges from 1 (not at all for health and wellness) to 7 (completely for health and wellness), with higher numbers representing a greater amount of motivation for health and wellness. Reported on the Wellness Management Technology Background Questionnaire.

Usage

Participants were experienced users. Twenty-eight participants (93%) used their tracker seven days per week. One participant used her tracker six days a week, and one participant used her tracker five days a week. Participants typically checked their activity trackers frequently, with 20 participants (67%) checking their activity tracker at least once every four hours. Twenty-five participants (83%) had been using their current activity tracker for three months or more. Of those 25 participants, 14 had been using their current activity tracker 6 months or more. When asked about how often participants used specific features and functions on their activity trackers, 29 participants (97%) responded "the automated wearable tracking device (e.g., the Jawbone Up bracelet) was "used frequently," 26 participants (87%) responded "the step tracker (e.g., steps walked/ran) was "used frequently," 24 participants (80%) responded "the phone app" was "used frequently" and 20 participants (67%) responded "the distance tracker" was "used frequently."

Perceptions of Activity Trackers

Participants found their trackers to be accurate between 61% and 100% of the time. We specifically assessed four features: step counting, duration asleep, distance moved, and calories burnt. Participants found the distance tracker to be the most accurate and consistent feature, and the sleep tracker to be the most inaccurate and inconsistent feature. Table 7 shows the number of boxes checked on the Automation Measurement Profile aggregated across all participants (where each box represents a 5% difference from the value the activity tracker provided). The boxes checked demonstrate the strength (i.e., 0% off, 5 % off, 10% off, or 15% off) and direction (i.e., overestimation or underestimation) of the perceived inaccuracy score for that feature.

In sum, participants were healthy and active with a moderate amount of technology experience. Participants were unsuspicious of automation. They reported being experience

Table 7
Totals of All Answer Choices Selected on the Automation Measurement Profile [a]

1. Your activity tracker reports you have walked 7,000 steps.	5,950 steps	6,300 steps	6,650 steps	7,000 steps	7,350 steps	7,700 steps	8,050 steps
	0	5	20	23	14	1	0
2. Your activity tracker reports you have slept for 8 hours.	6 hrs 48 min	7 hrs 12 min	7hrs 36 min	8hrs 0 min	8 hrs 24 min	8 hrs 48 min	9 hrs 12 min
	4	9	20	25	10	1	2
3. Your activity tracker reports you have moved (walked/ran) 4 miles.	3.4 miles	3.6 miles	3.8 miles	4 miles	4.2 miles	4.4 miles	4.6 miles
	0	5	20	25	17	6	1
4. Your activity tracker reports you have burnt 1,500 calories today.	1275 cal.	1350 cal.	1425 cal.	1,500 cal.	1575 cal.	1650 cal.	1725 cal.
	1	6	17	23	19	10	4

[a] The number provided below in each potential reading is the number of participants who responded that that reading could be the actual amount of the activity compared to what their activity trackers stated was the measurement. Participants could select more than one response option per question.

users, with almost all participants using their activity trackers daily and with most using their activity tracker for more than 3 months. About half of participants had additional experiences with other wellness management technologies to draw on when understanding automation errors. As indicated by the Automation Measurement Profile, participants generally found their activity tracker to be accurate within 5% across multiple functions. However, the interview data reveal what information participants use, and what they do, when the activity tracker performs inaccurately. We turn to the interview data next

CHAPTER 4

THREAT STRATEGY INTERVIEW RESULTS

The results of Threat-Strategy Interview (TSI) were examined to answer four research questions related to troubleshooting attend to automation issues:

- [R1] What information serves as a cue to an automation issue for experienced everyday automation users?
- [R2] With what level of detail do experienced everyday automation users interpret automation issues?
- [R3] Upon attending to an automation issue, how do experienced everyday automation users decide to respond?
- [R4] What strategies do experienced everyday automation users have for responding to an automation issue?

Participants were asked to describe an automation issue that could make it difficult for them to user their activity trackers. A total of 84 incidents was generated. Each participant self-segmented incidents based on their own perceptions. For example, if one participant thought that the distance over-estimating and the sleep tracker underestimating sleep time were really the same issue (e.g., a sensitivity problem), then the participant might talk about them as one issue. However, if a different participant thought those were separate issue (e.g., a stride-length error and a sensitivity problem), then the issues would be talked about separately. On average, participants provided 2.8 incidents (*SD* =1.13). However, the range varied substantially from 1 to 5. Most frequently, participants reported either 2 (n=10) or 3 (n=10) incidents.

The incidents generated by participants provided data regarding what participants thought was the issue and why they thought that was the issue, as well as what they would do about the issue and why they would do it. After the interviews were transcribed verbatim, issues and cues to issues were segmented into the same document and reactions and reasons for those reactions were segmented into a second document. This was done to provide context during coding (e.g., for when pronouns were given), and entire interview scripts were available during coding. Additional details regarding data segmentation and coding schemes are provided in their relevant sections below.

Automation Issues and Cues to Automation Issues

To examine [R1] and [R2] the transcripts to the following interview questions were segmented:

1. Can you tell more about the nature of the issue {or describe which issue you are talking about}?
2. How do you become aware of the issue?
3. If *error is specific*: Why did you think that [*repeat the cue(s) they mentioned*] was related to [specific problem]? OR If *error is vague*: Why did you think that [*repeat the cue(s) they mentioned*] was a problem with your {activity tracker}?
4. When did this issue occur?
5. Can you tell me what happened just before this issue occurred?

A cue to an issue was operationalized as "any non-repetitive thing (e.g., sign, something in the environment) that leads the participant to notice or understand the situation or error." An issue interpretation detail was operationalized as "the types of facets of knowledge expressed in describing the source or entirety or nature of the issue, such as why, how, or where the issue exists. Issue interpretation details should be non-repetitive." Thus, for each incident, it was possible to have more than one interpretation of what was the issue. Also, note that although cues were non-repetitive within a detailed error interpretation, the same cue may be used across multiple interpretations of the same error. In other words, participants could reinterpret their same incident multiple times and could use the same cues for all of those reinterpretations.

Issue Interpretation: [R2] With What Level of Detail Do Experienced Everyday Automation Users Interpret Automation Issues?

For issue details, a multi-level coding scheme was developed. According to Consolvo et al. (2008), many users do not consider it to be an error when an activity tracker fails to detect an activity that the tracker was not designed or trained to be able to detect. Thus, the first level of coding was either an automation error or not an automation error. Not an Error also applied if the participant felt the issue was a human error. Of those coded an automation error, the next division was between General and Specific Errors to capture the differences between automaton studies that provide cues to errors and those that simply set reliability to a certain percentage and allow errors to occur randomly (e.g., Riley, 1996). General errors occurred when participants acknowledged that something was wrong, but could not specify what was wrong or why it was wrong. Specific errors were coded at one final level, wherein they were categorized as Specific Non-causal or Specific-Causal. This division was also top-down and designed to capture the distinction in the human-automation literature wherein causal information is sometimes explicitly given to participants (e.g., Dzindolet et al., 2003) compared to times when just cues are given to participants (Itoh et al., 1999). Table 8 provides additional details of the automation issue coding scheme. The coding scheme underwent minor revision until an inter-rater reliability between two coders greater than 80% was reached both across participants and within incidents.

Table 8

Issue Interpretation Detail's Coding Scheme

Issue Detail Code			**Definition**	**Exemplar Quote**
Not a Automation Error			The participant explains that the device is supposed to work that way and thus it is not an error or the participant explains that the user made a mistake**.**	"I guess I assume that either I accidentally hit my Fitbit and turned it on by myself"
Automation Error	General Error		The participant states that an error, problem, issue, or other general term occurred. The participant does not articulate what exactly the error is or where it occurred. This might include, but are not limited to, situations where the participant explains that the error is ongoing or the error somehow makes the tracker inaccurate or reliable (e.g. over-estimation, under-estimation).	"It was an actual malfunction with the activity tracker"
	Specific Error	Non-Causal Error	The participant explains *where* or *what specific* feature or function erred. These are not limited to, but might include the distance, step counter, stair counter, sleep tracker, progress chart, calories, synching, or other specific feature or function. If what, the participant could explain that the issue ranges across all functions (e.g., a calibration issue).	"Sometimes if I haven't synced it {the tracker} for a while it will have a hard time finding it {my phone}"
		Causal Error	The participant explains how the error occurred or what caused the error. These may include a variety of specific errors with a logical explanation, information about how the device functions, damage to the entire tracker, and many other explanations. It typically includes a *how or why*, not just a *where* or *what.*	"There's a conversion from the number of steps you take to the reported number of miles you walk. I don't think it's very accurate because it doesn't take into account your stride"

From the 84 incidents, a total of 121 issue details were segmented and coded. Table 9 describes how many of these details occurred in each incident and the mean per participant. To start the analysis, the issue details were divided into Automation Errors and Not Automation Errors (Figure 3a). Most of the interpretation details (87%) in the incidents generated by participants matched the definition of automation error. A chi-square goodness of fit test demonstrated that Automation Errors details were overrepresented compared to Not an Automation Error, χ^2 (1, N = 121) = 68.44, $p < .001$. Participants with this code typically were uncertain if human error might be the reason for the issue, or if the issue was even real, as illustrated by this quote "*well I guess I don't know, because you're…sleep is a weird thing, so it's like 'well maybe I was awake?' I don't know*." Indeed, rarely did participants reference the programming limitations of the device, although participants did occasionally mention device properties with user error (e.g., "*so I recognized it was basically my fault because they indicate that you shouldn't get it wet beyond washing your hands or something*").

Table 9
Error Interpretation Details and Error Cues Generated by Participants

	Incidents	**Error Details**	**Cues**
Total	84	121	454
Mean per participant	2.8	4.03	15
SD per participant	1.13	1.95	8.2
Range per participant	1-5	1-9	4-36
Mean per incident	--	1.44	5.4
SD per incident	--	.82	3.3
Range per incident	--	--	1-25
Mean per detail	--	--	3.75
SD per detail	--	--	1.89
Range per detail	--	--	0-10

Of the automation error details, few (10%) were interpreted as General Automation Errors (Figure 3b). Rather, most of the details were specific, χ^2 (1, N = 106) = 66.57, $p < .001$. The General Automation Errors typically explained that the participants recognized that there was an automation error, but that they were unable to explain what was the error, as illustrated by these quotes: "*there was no real explanation, there was no information about why that might be*," "*I don't know technically why it was doing it*," and "*I'm not sure why it would overestimate it so much.*"

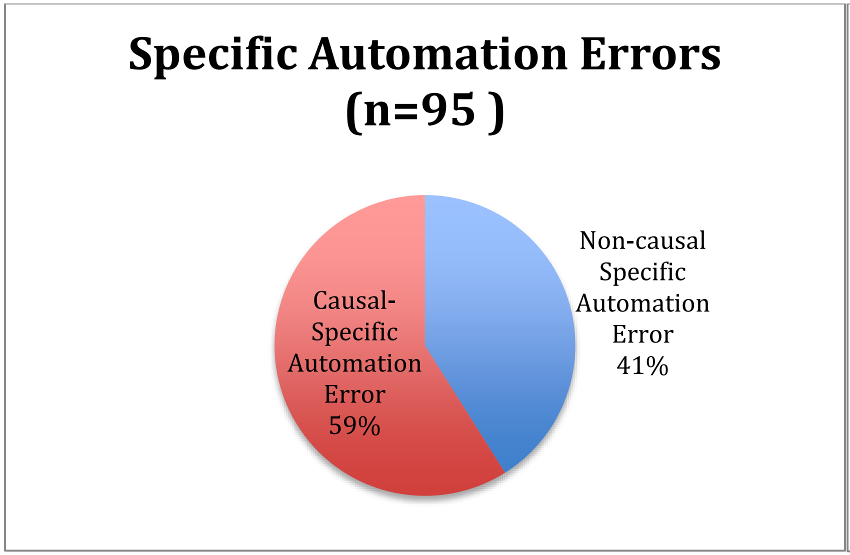


Figure 3a. Not an automation error and automation errors.

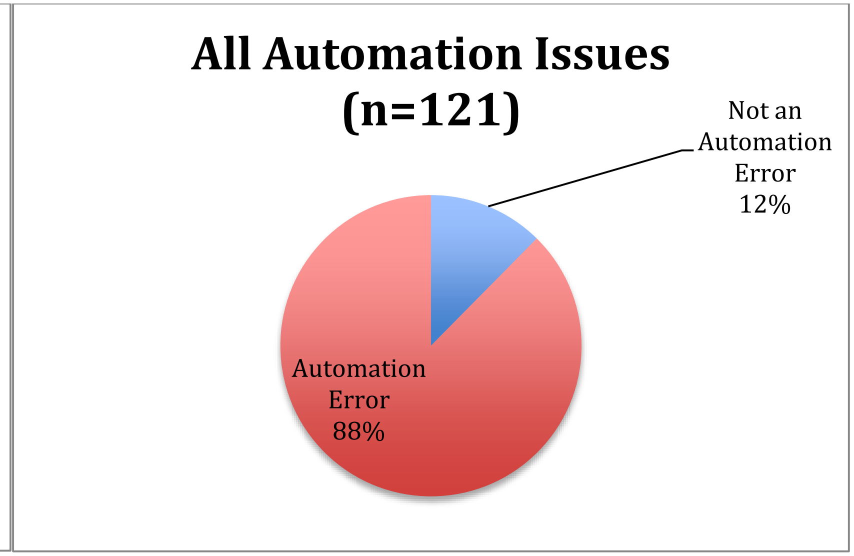


Figure 3b. General specific automation errors

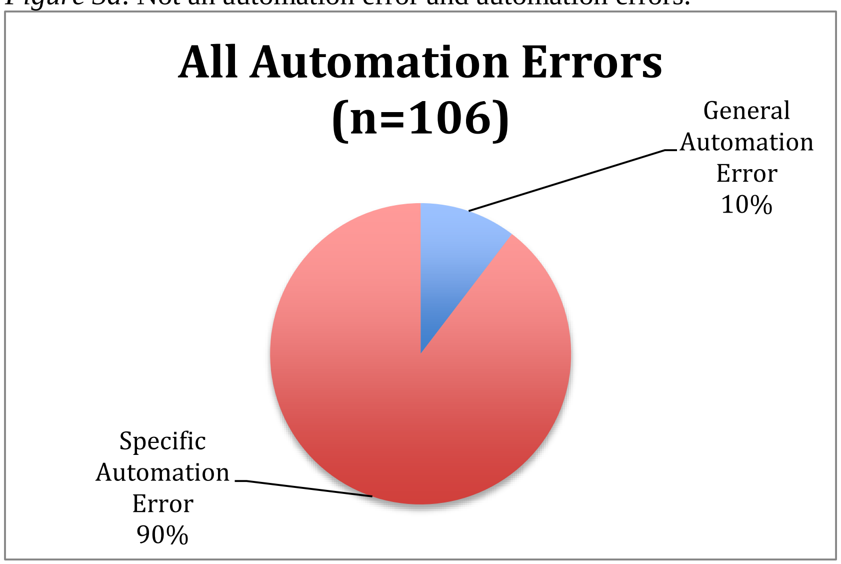


Figure 3c. Non-causal specific and causal specific automation errors.

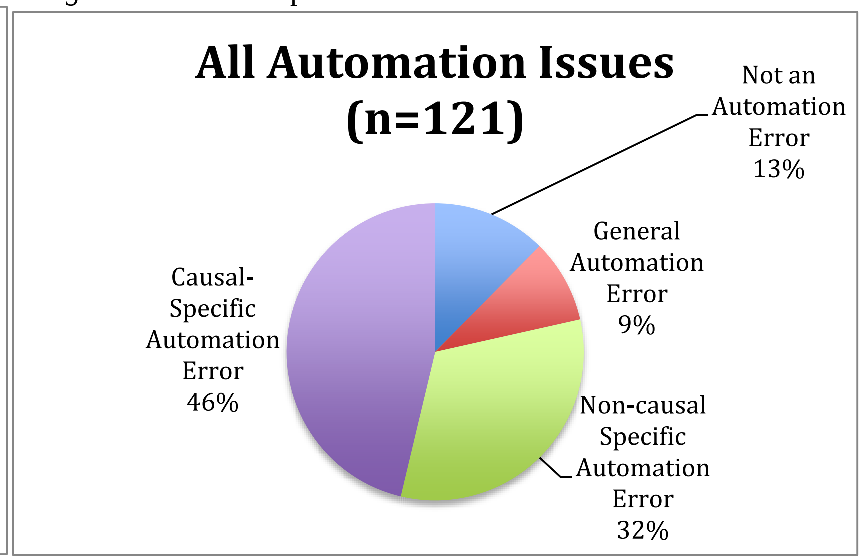


Figure 3d. All automation error codes.

Figure 3. Automation issue interpretation detail codes applied to the TSI data.

For the Specific Error detail interpretations, most of the explanations were Causal Specific as opposed to Non-causal Specific error details (Figure 3c), but this difference was not significant χ^2 (1, N = 95) = 3.04, *p* <. 081. Non-causal Specific Automation Error details generally involved explaining than the error was or was not confined to certain features or to certain situations. For instance, one participant explained, *"Yeah it's just running tends to underestimate the amount of steps that I take. Then, as a result, it tends to underestimate the amount of calories."* Non-causal Specific data were typically very detailed accounts of the situation, but lacked an explanation as to why the error was occurring, as was the case for this participant, "*The tracker was tracking its data as it should, but whenever I would try syncing it, it would start syncing and show progress but then it would just say 'failure to sync.'"* Similarly, Causal Error interpretations were typically very detailed, but in contrast to Non-causal Specific Error interpretations, Causal interpretations explained why the error was occurring. These quotes provide good examples of the nature of Causal Error explanations:

- *"It must have been that each step the horse took, the [Activity Tracker] mistook it for several human steps"*
- *"Or sometimes, I just move too much that it thinks that it's walking"*
- *"I've noticed that I've had times where it says that I didn't fall asleep until much later than when I actually did….But I think, I think that it tracks your movements like if you're not moving a lot, if your heart rate lowers I think that's when it assumes you're asleep....it'll say I have 50 steps or more steps than what I actually have. But I think it's tracking me moving in bed."*
- *"It's not taking into account that my heart rate is way higher when I'm burning those calories."*

However, in comparing the Causal details with Non-causal Specific, General, and Not an Error details, it becomes clear that although the most frequently mentioned type of interpretation (46%) was Causal in nature, Causal interpretations were still not the majority of all issue details (Figure 3d). A chi-square goodness of fit test demonstrated that all 4 groups were different from one another χ^2 (3, N = 121) = 44.39, $p < .001$. In sum, Causal Automation Errors were the most common interpretation, however other error interpretations, such as General and Non-causal Specific still occurred for experienced automation users.

Cues to Automation Issues: [R1] What Information Serves as a Cue to an Automation Issue for Experienced Everyday Automation Users?

The Issue Interpretation analysis determined what was interpreted. The following analysis of the cues to automation issues will help explain how those interpretations came to be, as in, what information was used to arrive at the interpretations of Not an Automation Error, General Automation Error, Non-causal Specific Automation Error, and Causal Automation Error.

Both top-down and bottom-up approaches were used to develop a coding scheme for cues to an automation issue. The top-down parts were developed from findings in the human-automation literature, the person perception literature, and the bottom-up parts were derived from material development and pilot testing. Table 10 lists the codes and the relevant citations, along with their definitions, and example of quotes from participants. The coding scheme underwent minor revision until an inter-rater reliability greater than 80% was reached both across participants and within incidents.

A total of 454 cues to issues were segmented and coded. The mean number of cues used for

Table 10

Cues to Automation Issues

Code	Definition	Citation(s)	Exemplar Quote
Context	Participant describes the context or a specific stimuli that the error does or does not occur in or around, this could include the task of the activity tracker being easy or difficult, or comparing it to other situations that other people have experienced (e.g., online forums)	Madhavan et al., 2006; Kelley 1967; Pop 2013	"I think mostly {during} travel -- I don't think it {the error} happens much if I'm just sitting at a desk…"
Measurement Comparison	Participant compares the activity tracker to other technologies or people. For example, it may be that other technologies or people may make similar errors, or it could be the readings provided by other information- such as a person's memory of their own behavior or a reference point such as a known distance or time	Parasuraman & Riley, 1997; Kelley 1967; Pop 2013; Masalonis, 2013	"Checking I guess on another pedometer like my phone"
Device Mental Model (i.e., Logic of Error)	Participant describes that the error should have occurred based on how the technology works, that the participant understands or comprehends why an error occurred that is related to how in functions, or otherwise states the error makes sense or mistook a similar stimuli as a target stimuli	Dzindolet et al., 2003; Lees & Lee, 2007; Bagheri & Jamieson, 2004; Bisantz & Seong, 2001; Norman, 2002; Van der Veer & Melguizo)	"Just sort of thinking about how if I were to design this, how…what does it track normally and how can somebody recreate that without actually doing the task."

Table 10 (continued)

Code	Definition	Citation(s)	Exemplar Quote
Checking Device	Participant does not cite reliability, distinctiveness or consistency from above. Rather the participant talks about checking their tracker and noticing something unexpected in general	Materials development/ pilot testing	"Just when I get my daily first thing in the morning notification, it's honestly one of the first things I look at when I get my coffee and I saw there was nothing there."
Consistency	Participant describes or questions the consistency or frequency of the error or the error repeating. This would include the distributions of errors over time and how much the device deviates by in its errors changes (i.e., amount of deviation variability)	Itoh et al., 1999; Muir & Moray, 1996; Kelley 1967; Pop 2013	"I become aware of it just be doing it multiple times, cause, like, sometimes, like, weird things will happen once, but when they happen over and over again."
Component Information	Participant describes that the information provided to the automation (either the activity tracker as a whole or calculations or other sub functions of the activity tracker) have been provided in/adequate information. The participant may describe other parts of the technology as experiencing similar or different problems, and may talk about it feeding through	Rovira et al., 2014	"So it's only basing it on average statistics, I'm sure, of someone my weight and height."

Table 10 (continued)

Code	Definition	Citation(s)	Exemplar Quote
Information Provide About Device	Participant describes reading/watching videos about their activity trackers and how or when errors can occur that the participant explains includes a cause (e.g., FAQs, user manuals, online forums prior to the error occurring) The idea is that the participant is told this may potentially be an error.	Dzindolet et al., 2003; Lees & Lee, 2007; Bagheri & Jamieson, 2004; Bisantz & Seong, 2001	"Well they tell you not to get it wet, so I imagine that if you do get it wet it will probably cause damage to it because otherwise why would they tell you to not get it wet."
Knowledge about Fitness or Health	Participant describes knowing how many calories a food contains, number of calories burnt in exercise, etc.	Materials development/ pilot testing	"So it gives you a base metabolic rate and standard calories burnt per time walked. So what I've learnt is that, um, just from online is that everyone has like a different build, even if you're the same weight and height. So two people with different builds could have different calorie outputs"
Testing out the Tracker	Participant describes knowledge gained by exploring their activity tracker, running pseudo-experiments, or actually doing the activity and discovering the error. As opposed to just checking the tracker, it's a	Materials development/ pilot testing	"And then I just ran an experiment. That's essentially what I did. And I, as we started, I showed my wife the stair count, and I started to swing my arms as I went up, and I showed it when I got off {the escalator}"

Table 10 (continued)

Code	Definition	Citation(s)	Exemplar Quote
General Sense or Feeling	Participant describes having a general hunch, or how tired/energized he or she feels, or may use some other general "seems like" or "feels like" term.	Materials development/ pilot testing	"I think maybe possibly is just the fact that it's unknown to me that I don't trust it"
Other	Cue does not fall into any of the above categories.	--	"That's really the only explanation I can come up"

issue error interpretation was 3.75. In 2 interpretations, no cues were provided, although in each case there were multiple interpretations for the incident and the other interpretations did code was not equivalent across each type of issue interpretation, χ^2 (18, n = 402) = 36.16, p = .007. In other words, cues and issue interpretations are indeed related have cues. Thirteen cues were eliminated from further analyses because they did not fit into any category under the coding scheme. Of the remaining ten groups, a χ2 goodness of fit test showed that not all cues were reported equally, χ2 (9, N = 441) = 154.39, $p < .001$. Context (n=82) and Measurement Comparison (n=80) were the most frequently mentioned cues (Figure 4). Three of the most underrepresented categories of cues were Knowledge about Fitness or Health, General Sense or Feeling, and Testing out the Tracker. Each of these 3 categories made up less than 5% of the total cues. Additionally these three cues were not represented across all the types error interpretations. To compare the types of cues to errors across different types error interpretations, these three categories were also eliminated from further analyses. Table 11 reports the remaining 402 cues across the remaining 7 categories. Even with the three most underrepresented categories eliminated, a χ2 goodness of fit test showed that not all cues were reported equally, χ2 (6, N = 402) = 44.71, $p < .001$. As demonstrated in Table 4, Information Provided About the Device (n=26) and Component Information (n=37) were reported less frequently than other cues. Additionally, a χ2 test for independence revealed that the portion of each cue

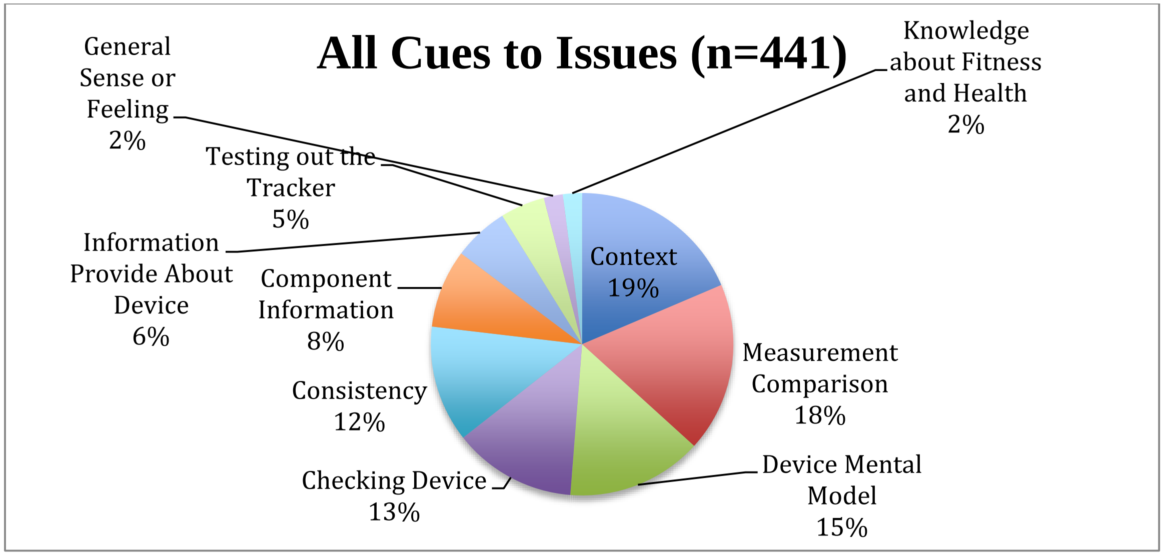


Figure 4. All the cues to issues, except for those that fell into the "Other" code, from the TSI data.

Table 11

Percentages of Cues for Each Category of Issue Interpretation

Category	**Total (n[1]=402)**	**Causal Specific Automation Errors (n=201)**	**Non-casual Specific Automation Errors (n=121)**	**General Automation Error (n=27)**	**Not an Automation Error (n=53)**	**All Automation Errors (n=349)**	**All Specific Automation Errors (n=322)**
Context	20	26	17	26	6	23	22
Measurement Comparison	20	16	22	19	30	18	18
Device Mental Model	16	18	9	7	28	14	15
Checking Device	14	12	17	15	17	14	14
Consistency	14	11	18	22	11	14	13
Component Information	9	11	9	7	2	10	11
Information Provide About Device	7	6	8	4	6	7	7

[1] n is the number of cues within each type of error interpretation

code was not equivalent across each type of issue interpretation, χ^2 (18, n = 402) = 36.163, *p* = .007. In other words, cues and issue interpretations are indeed related.

Descriptively, and as Table 11 shows, in determining if an issue is an Automation Error or Not an Automation Error, Measurement Comparison and Device Mental Model cues helped users reach the conclusion that the issue was Not an Automation Error. Context and Component Information were more frequent for issues participants found to be an Automation Error compared to Not an Automation Error. For those issues deemed an Automation Error, Consistency was a comparatively frequent cue for General Automation Error interpretations, whereas Information Provide About the Device was more frequent for Specific Errors. Within Specific Errors, Causal interpretations were heavily associated with Context cues, and Device Mental Model cues whereas Non-Causal Specific Automation Error interpretations were more associated with Measurement Comparison, Consistency, and Checking Device cues. In sum, many cues are used in interpreting an automation issue, but the cues used are related to the interpretation to which the user arrives.

Strategies for Responding to Automation Issues and Reasons for Strategy Selection

To examine [R3] and [R4] the transcripts to the following interview questions were segmented:

1. Can you tell me a strategy you might use to keep that issue {or describe the issue} from interfering with tracking your activity, fitness, or heath?
2. Can you tell me why you would try that particular strategy {or, if multiple strategies, describe the ones you are talking about}?
3. *Only if they have an intervening strategy:* How can you tell if {repeat actions or strategy A} is working or is not working?

4. What would you do if {repeat actions or strategy A/that} did not fix the problem?
5. Why would you choose that particular strategy {strategy B}?

   Stopping rule: discontinue use permanently, not to track, to track wrong, to send back to manufacture, to manually override the automation, or to track wrong and correct in your mind.
6. Did this situation actually occur?
   a. *If yes,* what did you actually do?
   b. *If yes,* what was the outcome of what you actually did?
   c. *If yes,* did you do anything else in response? Maybe later on?
7. Now that we have talked about all the strategies you actually used or could have used, I would like to know more about what would make you choose {the first actions or strategy A } over {the second actions or strategy B} or the other way around. Could you think about a specific piece of information or a cue that would encourage you to choose {strategy A} over {strategy B}? * Included for any strategies the actually did but didn't explain.
8. Were there any previous experiences you have had that was helpful to you in thinking about how you could deal with this situation?

A reason was operationalized as "any non-repetitive motivation (e.g., want, need) that may include cognitive or affective reasoning, for carrying out a particular behavior or a particular series of behaviors. These can include signs or something in the environment." Barrier reasons were also segmented and were operationalized as "any non-repetitive motivation (e.g., want, need) that may include cognitive or affective reasoning, for **not** carrying out a particular behavior or a particular series of behaviors. These can include signs or something in the environment." Additionally, if reasons were given for carrying out actions in general, as in without specifying a particular behavior, they were segmented as general reasons. Although reasons and barrier reasons were not included if they were repeated for the same action, reasons and barrier reasons could be repeated across the different actions generated by the participant in the same incident.

In accordance with TSI methodology, transcripts were first segmented for actions. An action was operationalized as "any non-repetitive instance in which a person carries out or could (if hypothetical) carry out one or more behaviors with the goal of alleviating the issue in a distinct or unrelated way from other behaviors." Any example of one behavior would be to do nothing and an example of an action with more than one behavior would be to search for help online and follow online instructions that suggest restarting the activity tracker (i.e., in this example, restarting the tracker is dependent upon searching online). Each action for each incident for each participant was first segmented. Then, two raters grouped each action with other actions from the same incident (and participant) that were non-repetitive but very similar. If both raters grouped actions together, then those remained as an action set for the duration of the analysis. However, these groupings were rare and instead most action sets contained a single action and were thus single-item action sets. In total, there were 4 two-action sets, and 273 single-action sets.

Strategies for Responding to an Automation Issue: [R4] What Strategies do Experienced Everyday Automation Users Have for Responding to an Automation Issue?

As per TSI methodology, the strategy coding scheme was data-driven. The segmented action sets were combined across all incidents and all participants and sorted again into groups of similar action sets (Gregg, Ferguson, & Durso, 2014). Two independent judges did this second sort. Inter-rater agreement was calculated according to TSI methodology and was 95.6% for this second sort. Table 12 provides an excerpt from this second sort of action sets and shows what counted as agreements between the judges. All the actions sets were re-arranged to a mini-max rule wherein groups covered the greatest number of sub-groups (i.e., sorted together by one rater, but not the second rater) while excluding the least amount of sub-groups. The original 49 groups were immediately reclassified into 37 groups because of the extreme similarity between some groups (e.g., two of the 49 groups were both named by the coders “check for human error,” and the two “check for human error” groups contained very similar actions. Thus, those two “check for human error” groups were combined into one group.). The 37 groups of strategies, their names, and an example action set from each of the groups, are provided in Appendix N.

Table 12
Part of the Action-Set TSI Sort.[1]

Action Set	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	X	X	X	X	X	/	/	/	/	/	/			
2	X	X	X	X	X	/	/	/	/	/	/			
3	X	X	X	X	X	/	/	/	/	/	/			
4	X	X	X	X	X	/	/	/	/	/	/			
5	X	X	X	X	X	/	/	/	/	/	/			
6	/	/	/	/	/	X	X	X	X	X	/		\	\
7	/	/	/	/	/	X	X	X	X	X	/		\	\
8	/	/	/	/	/	X	X	X	X	X	/		\	\
9	/	/	/	/	/	X	X	X	X	X	/		\	\
10	/	/	/	/	/	X	X	X	X	X	/		\	\
11	/	/	/	/	/	/	/	/	/	/	X	\		
12											\	X		
13						\	\	\	\	\			X	\
14						\	\	\	\	\			\	X

[1]Each number represents an action set. One rater’s groupings are represented by “/” and the second rater’s groupings are represented by “\” An “X” represents agreed upon groupings by both raters. X’s represent positive agreements and blank cells represent negative agreements. Because each action set must be sorted with itself, the diagonal is composed of X’s. Shaded cells represent and instance where the first rater’s groupings were honored and the second rater’s groupings were not used. For example, rater 2 grouped 6,7,8,9,10,13, and 14 together, and rater 1 grouped 6,7,8,9,10,11,1,2,3,4, and 5 together. Whenever the two raters had different groupings, action sets were arranged to have the largest number of actions sets in a group and the fewest number of groups.

The 37 groups were then re-sorted into smaller groups by two judges. The first judge created the category names. Although the groupings were generally data driven, we included a Continue to Use category and a Change Usage Pattern category based on typical responses in human-automation studies. The second judge sorted also sorted the 37 groups into the categories, with a 81% reliability rating. Through discussion, category names were revised slightly and disagreements were resolved. The final six categories of strategies are presented here. These categories of strategies were:

1. Continue to Use
2. Change Usage Pattern (e.g., only use for some activities, use with a back up, disuse, use at different times)
3. Gather Information or Seek Help to Get it Fixed
4. Wait for Something to Happen
5. Change or Monitor My Behavior in the Situation
6. Try to Fix it on My Own.

The assignments of each of the 37 strategies into the 6 final categories are also presented in Appendix N.

From the 84 incidents generated by participants, 277 action sets emerged. For each incident, the most frequent number of strategy actions provided was 2. However, there was a considerable range, with incidents having up to 10 strategy actions provided (Table 9).

The most frequently reported type of action set (Figure 5) fell within the strategy category of Change Usage Pattern (n=95), followed by Continue to Use (n=54), and by Change or Monitor My Behavior in the Situation (n=54). A χ^2 test for proportions showed that these three categories were indeed over-represented and the categories of Gather Information or Seek Help to Get it Fixed, Wait for Something to Happen, and Try to Fix it on My Own were underrepresented, χ^2 (5, n = 277) =98.74, $p < .001$.

For the most frequent categories, Change Usage Pattern contained a variety of action sets that typically themed around not using the automation on the activity tracker, or using the

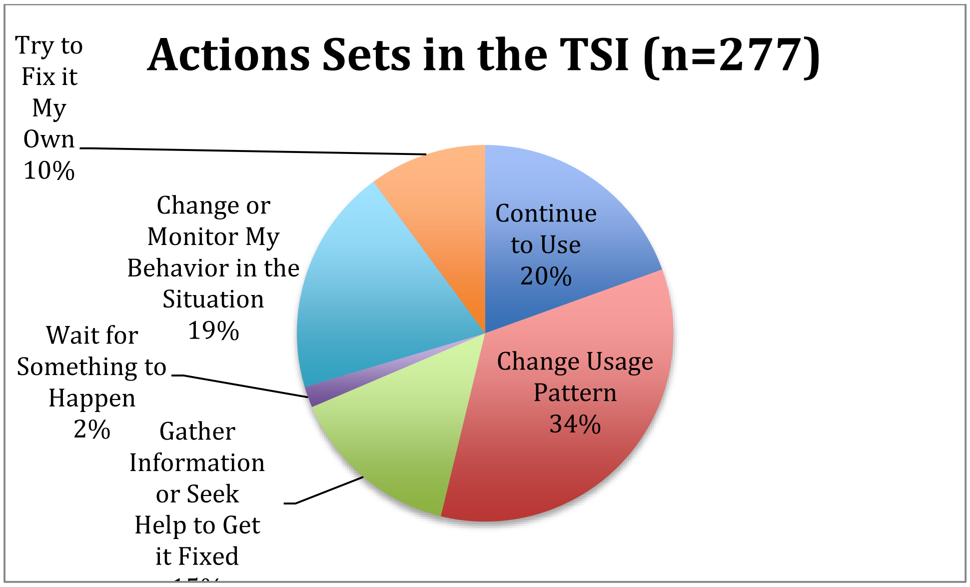


Figure 5. Action sets in the TSI categorized into strategy groups.

automation on the tracker in combination with a different method. Not using the automation included strategies like not using the activity tracker at all (e.g., "*the best way to do it would probably just not to use the….device itself,*" and *"I would stop doing it"*) but also actions like manually entering data into the activity tracker's app or website. Sometime participants described continuing to use the tracker, but just for certain activities, as described by this participant "*maybe only wear it when I'm working out.*" The Continue to Use category of strategies was largely composed of ignoring the automation issue actions (60% of Continue to Use). However, the category still contained many action sets that described correcting for the automation issue in the user's mind (40%), as this participant explained her action when she, *"just did the mental understanding it was going to be overestimated and just reversing it in my mind."* The Change or Monitor My Behavior in the Situation category of strategies was largely (50%) composed by actions that referenced changing behaviors to prevent the automation issue from happening as these two participants described "*"so I would try and keep it from getting wet"* and *"so I get the towel and try to like wipe off my arm before I do it and just keep it looser to where when I get home."* The remainder of the category included many responses about carrying out certain movements to make the activity tracker more accurate in the situation and about adjusting how the tracker is worn.

The less frequently reported categories of strategies included Gather Information or Seek Help to Get it Fixed, Wait for Something to Happen, and Try it Fix it on My Own. The Gather Information or Seek Help to Get if Fixed was largely composed of actions about looking for help online (46%) as this participant would, "*I look up the frequently asked questions on the Fitbit website and I follow that*" and contacting the company (27%) like this participant did, *"I emailed Fitbit customer support."* The Wait for Something to Happen category was fairly small (n=5) and included "waiting-out" the error or waiting for feedback from the activity tracker. Half of the Try to Fix it on My Own category contained general solutions such as restarting or resetting the tracker (e.g., "*I'd probably try to reset.").* The other half of the category included actions such as replacing part of the activity tracker and trying to synch the activity tracker with a different technology.

Reasons for Responses: [R3] Upon Attending to an Automation Issue, How do Experienced Everyday Automation Users Decide to Respond?

A coding scheme was developed for the reasons participants gave for carrying-out or not carrying-out an action (e.g., look for information online) or series of action (e.g., look online and follow the direction online for troubleshooting and resetting the tracker). The coding scheme was developed from a mix of both top-down and bottom-up methodology. Some initial codes were based on the troubleshooting studies of Konradt (1995) and Bereiter and Miller (1989) wherein specialists used reasons such as ease if implementation, experience, past system failures, and causality or the certainty of the issue. Other top-down codes were based on the human-automation interaction literature, such as incorporating the extent of the issue's inaccuracy (i.e., how unreliable is the automation; e.g., Wickens & Dixon, 2007). Other initial codes were data-driven from materials development and pilot testing and included reasons related to the importance of the error, affect, cost, purpose of use, awareness, knowledge about the issue (e.g., the issue's location), and situational consistency (for both the user and other users). The initial coding scheme was revised until two independent coders had inter-rater agreement greater than 80% for at the incident level and across participants.

These codes were later combined into and put into a hierarchal scheme of Person Reasons, Situation Reasons, and Device Reasons. Within the Person Reasons, reasons were further classified into:

1. General Life Knowledge or Experience
2. Non-causal Knowledge or Experience with the Activity Tracker
3. Figuring Out What Caused the Issue
4. Awareness
5. Personal Preference
6. Importance of the Issue/Purpose of Use.

Situational Reasons included:

1. Cost

2. Ease of action implementation
3. The availability (or lack thereof) of other alternative strategies (No Alternative).

The Device Reasons category include the Extent of the Automation Issue and Situational Consistency. The detailed coding scheme is described in Table 13. The same coding scheme was utilized for all reasons (i.e., reasons to do an action, reasons not to do an action, and general reasons for unspecified actions). Less than 5% of all reasons did not fit into any of the coding scheme and were eliminated from the further analyses presented here.

Two of the 277 action sets did not include any reasons. The remaining action sets were associated with 12 general reasons, 91 barrier reasons, and 647 reasons to carry out an action. Nine out of thirty participants provided a general reason. Twenty-nine participants provided at least one barrier reason, and the most frequently reported number of barrier reasons per incident was 1 (n=37), with a mean number of barrier reasons per incident of 1.4 (range= 0-6, SD=1.2). However, 8 action sets only contained barrier reasons (and no reasons for carrying out an action). All incidents (and all participants) provided reasons for carrying out strategies, with a mode of 4 (n=13) reasons per incident and a mode of 10 per participant (n=3;Table 14).

Most (83%) of the general reasons were Person Reasons, such as knowledge about other similar technologies, experience obtaining information, or knowledge about their own behavior. For example, one participant explained, *"Just the things that I know about other electronics have helped me troubleshoot this."* Because there were so few general reasons, they were not explored further.

The 91 barrier reasons were most frequently Person Barrier Reasons and Situation Reasons, and rarely were Device Reasons (Figure 6), χ^2 (2, n = 91) =19.07, $p < .001$. In looking deeper into Person Reasons, The Importance of the Issue and Purpose of Using the Activity Tracker, Personal Preferences, and various types of Knowledge and Experience were frequently reported (Figure 7). In examining the Situation Barrier Reasons, most of the reasons involved the Ease of Action Implementation (Figure 8). Most of the Device Reasons related to the Extent of the Issue (e.g., how much the automation was off by or how frequently the issue occurred), as seen in Figure 9. Because cell sizes were smaller than 5, χ^2 tests for proportions were not conducted.

Table 13
Reason Codes for actions used in the TSI and SBI

High Level Reason Code	**Low Level Reason Code**	**Definition**
Person Reason	General Life Knowledge/Experience	This code includes statements demonstrating knowledge experiences other than those with an activity tracker. For example, participants might talk about experiences with their own behavior or they might talk about other technologies (e.g., compare their activity tracker to a phone). The also could compare their activity to other errors phones or people make. Or the participant might talk about experiences obtaining information. Keep in mind that this code could also include references to a lack of knowledge or lack of experience.
Person Reason	Non-causal Knowledge/Experience about the Activity Tracker or Error	This code includes statements where the participant reference knowledge (or lack thereof) or experience (or lack thereof) with their current (or previous) activity tracker, but that knowledge *does not* explain *why* the error occurred. For example, the participant might explain knowing/not knowing exactly where the error is occurring as a primary reason for their decision. They might explain knowing the activity tracker has a website or an app where data can be entered.

Table 13 (continued)

High Level Reason Code	**Low Level Reason Code**	**Definition**
Person Reason	Figuring out What Caused the Issue	This code is similar to 102, but it includes a cause or an attempt to figure out a cause as the cues driving the participants response. Attempting to figure out a cause would include anything that relies on feedback that the error is or is to fixed after/before doing a certain response, or trying to eliminate possible causes. References to uncertainty may also be incorporated in this category, such as uncertainty as to what caused the error. Overall, this code also gets at knowledge (or lack there of) and experience (or lack there of) about activity trackers, but includes a cause. This code also applies if the user blames the issue on the user. This code applies to the participant describing his or her action (or lack thereof) as cued by the belief that the error is because of the user's fault. Alternatively, the opposite could also be true where the participant says the error is not his or her fault and that cues his or her actions. Another key word might be "human error." The comments should be explaining if the technology or the human is to blame.
Person Reason	Importance of Error/Purpose of Use	This code would include statements about why the participant uses the tracker or if the participant really cares very much or very little about the error.
Person Reason	Personal Preference	This code would include statements that talk about feelings, comforts, likes, dislikes, (and any other similar notions) as reasons for wanting to carry out or not carry out a particular response.
Person Reason	Awareness	This code includes the participant describing needing to be aware (or being unaware) of the error in order to respond. For example, the participant might describe reacting to the error only if he or she notices it.
Situation Reason	No Alterative	This code applies to statements in which the participant explains that he or she has no other option but a particular response. This might be because they do not know what else to do, or because there are no other alternatives.

Table 13 (continued)

High Level Reason Code	**Low Level Reason Code**	**Definition**
Situation Reason	Cost	The code includes monetary reasons as cues for a why a participant would or would not carry-out a certain response.
Situation Reason	Ease of implementation	This code would include statements about the effort needed to carry-out a certain reaction, the amount of time needed, and the overall convenience factor.
Device Reasons	Extent of the Issue	This code would include statements about the extent of inaccuracy (the deviation between the stated inaccurate value and the actual value) as a reason for determining to implement/not implement a strategy. It would also include information about the frequency at which the error occurs (e.g., the error happens a lot or the error happens rarely).
Device Reasons	Situational Consistency	This code includes comparing across or within situations (regardless of if those situations were experienced by the person or by others) in determining how to respond to the error. For example, the participant might want to know if other people are having the same issue (presumably in the same situation), or if there is a context or a stimuli where the error does or does not occur. This would also include asking people or looking online to see if the same situation had happened to other people.

Table 14
Number of Action Sets and Reasons in the TSI

	Incidents	**Action Sets**	**Reasons to carry out Action**
Total	84	277	647
Mean per participant	2.8	9.23	22
SD per participant	1.13	5.75	14
Range per participant	1-5	2-29	5-63
Mean per incident	--	3.30	7.7
SD per incident	--	1.81	4.5
Range per incident	--	1-10	1-25
Mean per action set	--	--	2.34
SD per action set	--	--	1.34
Range per action set	--	--	0-8

In examining how barrier reasons were related to strategies, the Wait for Something to Happen strategy group was excluded from a statistical test to enable adequate cell size. A χ^2 test determined that the number of barrier reasons in each strategy group was proportionate to the number of actions in each strategy group, χ^2 (4, n=90) = 6.94, p=.139 (Table 15). χ^2 tests were not conducted at either code level for each strategy group (Person, Situation, Device, or the more detailed coding scheme) because cell size was too small, even after excluding the Wait for Something to Happen strategy group. However, patterns are described descriptively here.

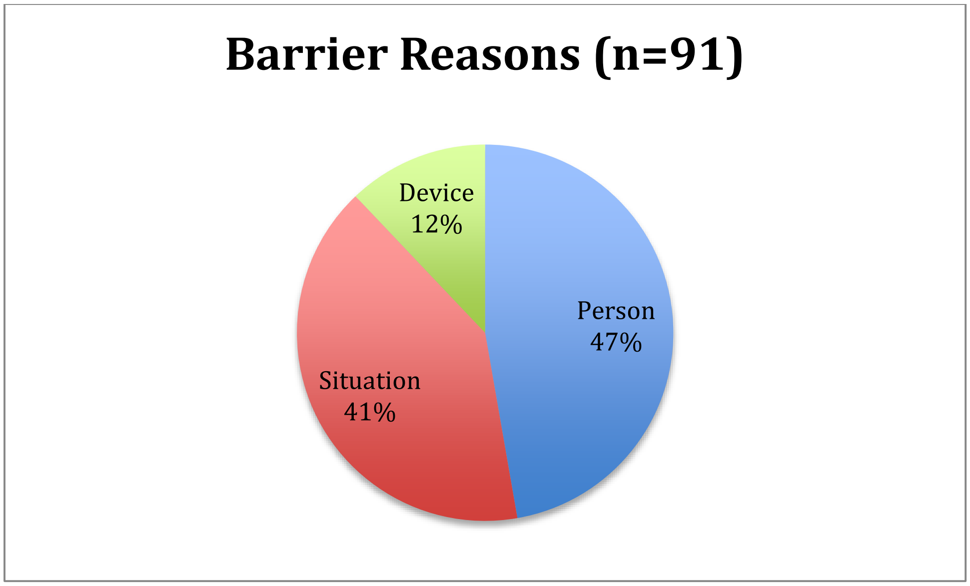


Figure 6. Reasons participants did not want to carry out a response in the TSI.

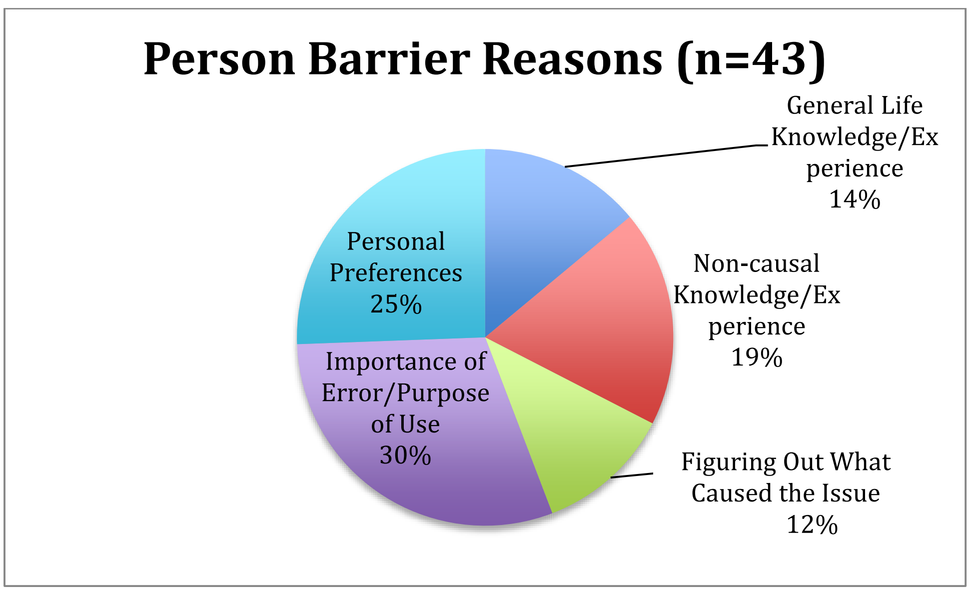


Figure 7. Person Barrier Reasons participants did not want to carry out a response in the TSI.

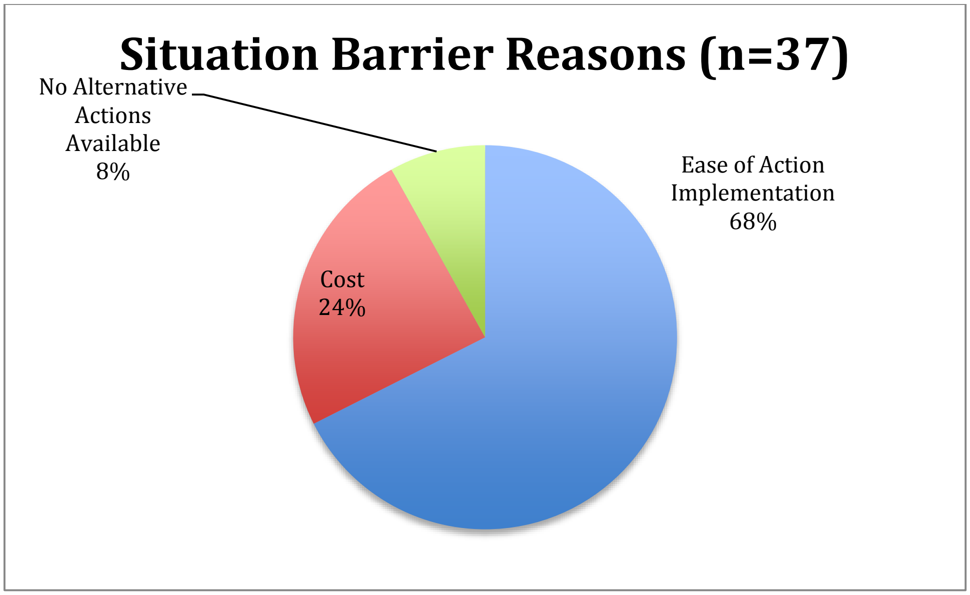


Figure 8. Situation Barrier Reasons participants did not want to carry out a response in the TSI.

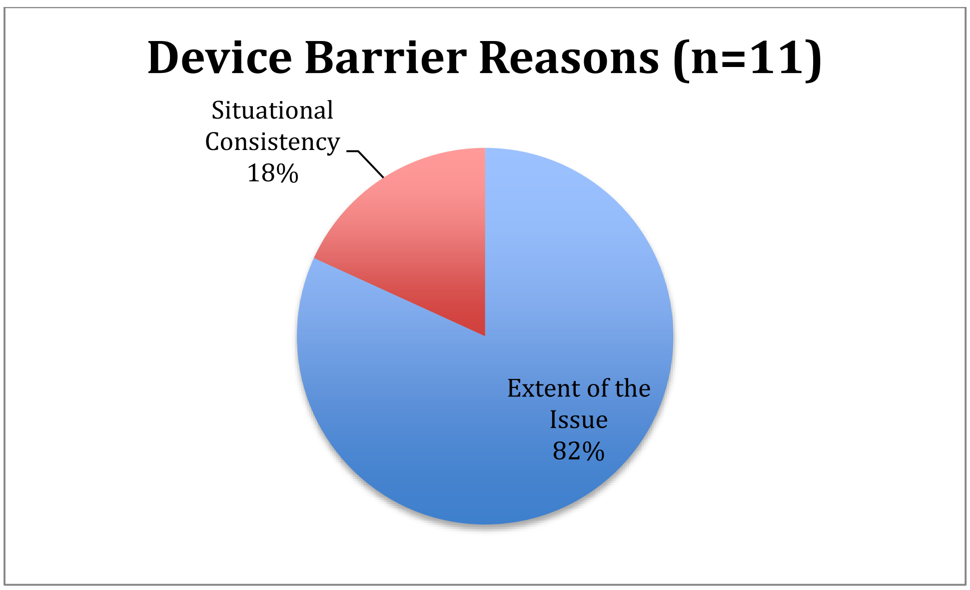


Figure 9. Device Barrier Reasons participants did not want to carry out a response in the TSI.

The most frequently reported strategy, Change Usage Pattern, also had the greatest number of barrier reasons (n=36). Almost half of the barriers to Change Usage Pattern were Person Reasons. For example, this participant explained she did not want to change her usage pattern to synching *"right after I've worked out to see if that will help it sync"* because *"in the work week*

Table 15

Barrier Reasons for Not Completing Actions in Strategy Groups[1]

	Continue to Use (n[2]=54)	Change Usage Pattern (n=95)	Gather Information or Seek Help to Get it Fix (n=41)	Change or Monitor My Behavior in the Situation (n=54)	Try to Fix it on My Own (n=28)	Total Of Each Barrier Reason Across Strategy Groups (n=272)
Person Barrier Reasons						
Figuring Out What Caused the Issue	0	1	1	2	1	5
Non-causal Knowledge/Experience	0	4	0	1	3	8
General Life Knowledge/Experience	1	1	1	3	0	6
Importance of Issue/Purpose of Use	3	8	1	0	1	13
Personal Preference	1	3	2	3	2	11
Awareness	0	0	0	0	0	0
Total Person Barrier Reasons	5	17	5	9	7	43
Situation Barrier Reasons						
No Alternatives	0	1	1	1	0	3
Ease of Implementation	2	7	5	6	4	24
Cost	0	5	2	0	2	9
Total Situation Barrier Reasons	2	13	8	7	6	37

Table 15 (continued)						
	Continue to Use (n^2=54)	Change Usage Pattern (n=95)	Gather Information or Seek Help to Get it Fix (n=41)	Change or Monitor My Behavior in the Situation (n=54)	Try to Fix it on My Own (n=28)	Total Of Each Barrier Reason Across Strategy Groups (n=272)
			Device Barrier Reasons			
Situational Consistency	0	1	0	1	0	2
Extent of the Issue	2	5	1	1	0	9
Total Device Barrier Reason	2	6	1	2	0	11
			Total Barrier Reasons			
Total of Barrier Reasons in Each Strategy Group	9	36	14	18	13	90

[1] This table excludes the Wait For Something to Happen Strategy Action group because the group had only 1 barrier reason and 5 action sets. Wait For Something to Happen is also not included in the total numbers.
[2] n is the number of action sets within each type of strategy group

*hustles {*a challenge feature*}, I don't want people to know how many steps I did, because it still gives them enough time to catch up."* As another example, a different participant explained that he did not want to return the activity tracker because he "*didn't want to send a gift back."* Change Usage Pattern also had the greatest amount of Situation Barrier Reasons compared to other strategies. One participant explained he might not "*try a different [ACTIVITY TRACKER] with a heart rate" "because that would cost some money."* Change Usage Pattern had the greatest number of Device Barrier Reasons compared to other strategies, although Device Barrier Reasons were less frequently mentioned than both Person and Situation Barrier Reasons. An example of a Device Reason was a participant who explained that she might not manually add the amount of calories she burned during exercise because her activity tracker might be underestimating (i.e., counting some, but not all, of her calories burned during exercise) and therefore she did not want to add the exercise manually and "double-dip"- *"you're more active in general you're still probably accumulating some steps, so you might double up on some of them. It's not a perfect fix."* For the remaining strategy types, Device Barrier Reasons were rarely mentioned.

Change or Monitor My Behavior in the Situation also had a substantial amount of barrier reasons (n=18). The types of barrier reasons typically mentioned for not changing the user's behavior were Person Barrier Reason (n=9) and Situation Barrier Reasons (n=7). For instance, this participant described the action of wearing her activity tracker elsewhere, "*I think I actually researched on this...I think you can wear it somewhere else*" but she also explained a Person Reason as to why she was weary to do so "*I don't want that accessory on me."* A different participant provided a Situation Barrier Reason for a similar action of wearing his tracker elsewhere - "*You can try putting it on the opposite arm*" but *"it's harder to read that way."* The strategy groups of Gather Information or Seek Help to Get it Fixed and Try to Fix it on My Own also had noticeable amounts of Barrier reasons that were mostly about the person or the situation. For instance, a Person Barrier Reason of *"I like to try to fix things myself before getting other people involved*" was provided for why a participant did not want to contact customer service and a Situation Barrier Reason of *"It seems like a lot of extra work*" was provided for a why participant did not want to plug his activity tracker into his charger and then the charger into his laptop to try to fix the problem on his own. Other strategy groups had few barrier reasons (less than 10).

The 647 reasons to carry out an action were mostly Person Reasons (n=436). However Situation Reasons and Device Reasons were also mentioned (Figure 10). A χ^2 test for proportions revealed that Person Reasons were significantly overrepresented compared to Situation and Device Reasons, χ^2 (2, n = 647) =341.94, p < .001. Diving deeper into Person Factors, the Importance of the Error or the Purpose of Use, along with various types of knowledge were frequently mentioned compared to Awareness and Personal Preference, χ^2 (5, n = 436) =126.46, p < .001 (Figure 11). In looking at just the Situation Reasons, it becomes clear that Ease of Implementation is a factor that is often considered when selecting a response to an automation issue. However, the availability of other alternative responses and monetary costs also contribute to the decision, although they contribute less frequently then Ease of Implementation, χ^2 (2, n = 127) =35.45, p < .001 (Figure 12). Lastly, in examining only Device Reasons, Situational Consistency and the Extent of the Issue were about equal in frequency, χ^2 (1, n = 84) =2.33, p = .127 (Figure 13).

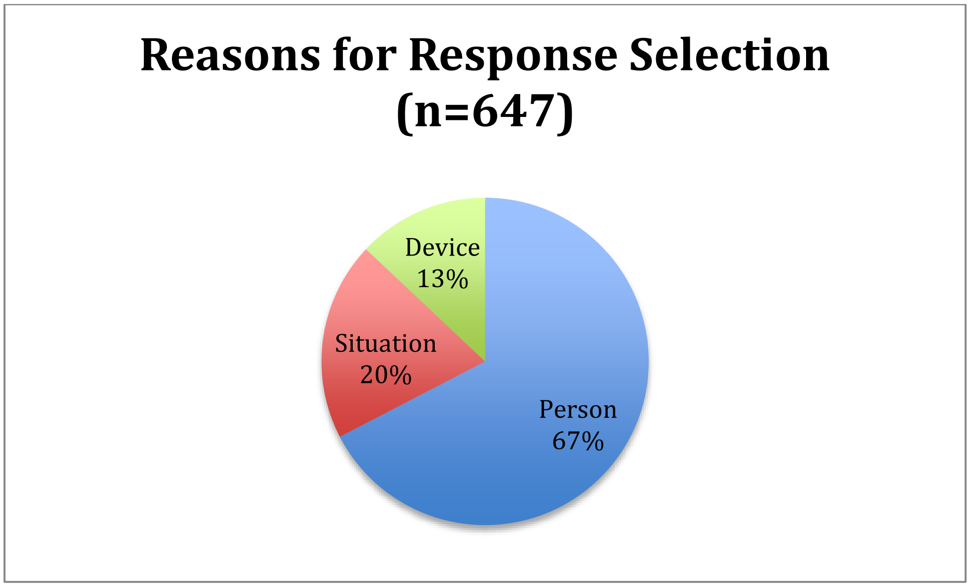


Figure 10. Reasons participants provided for carrying out a response in the TSI.

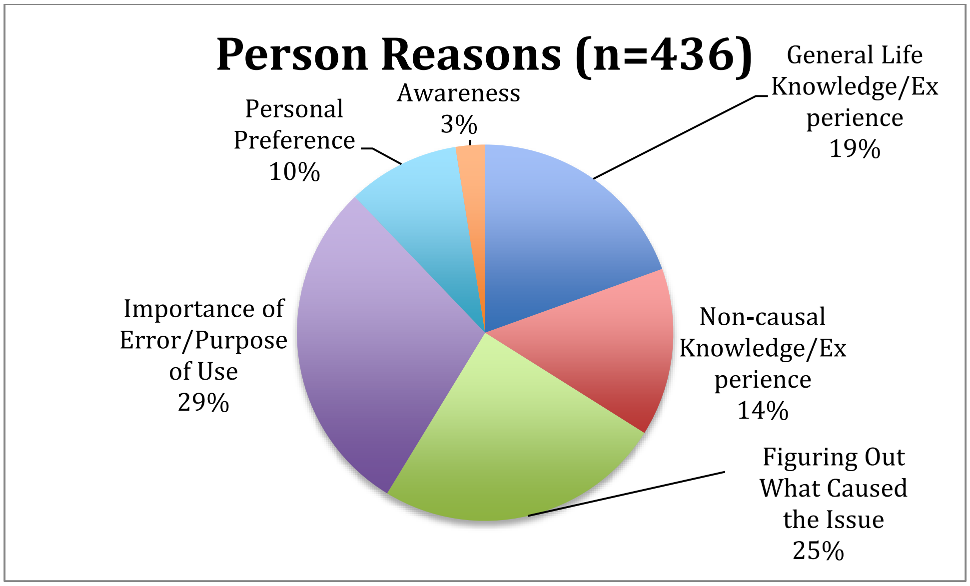


Figure 11. Person Reasons participants provided for carrying out a response in the TSI.

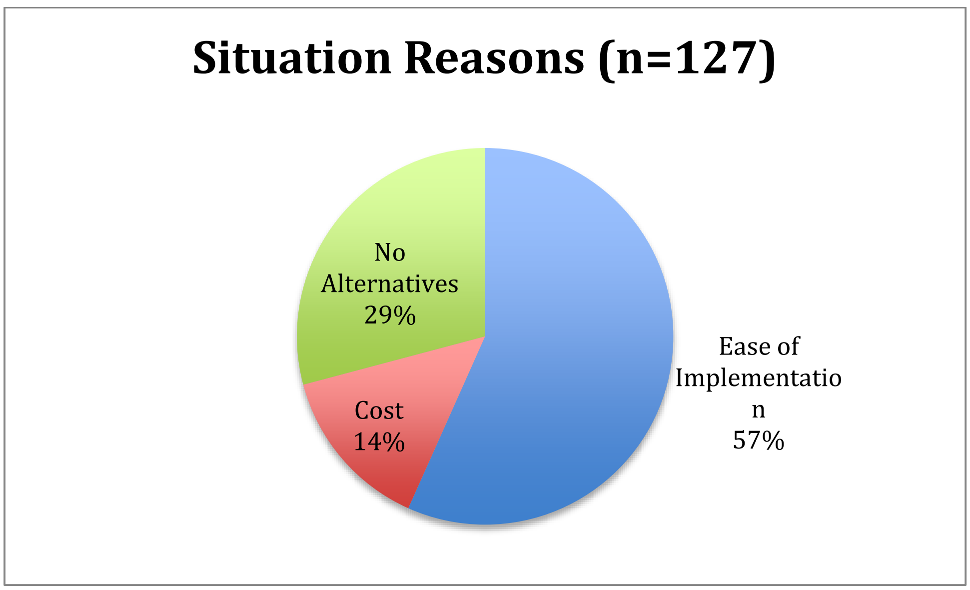


Figure 12. Situation Reasons participants provided for carrying out a response in the TSI.

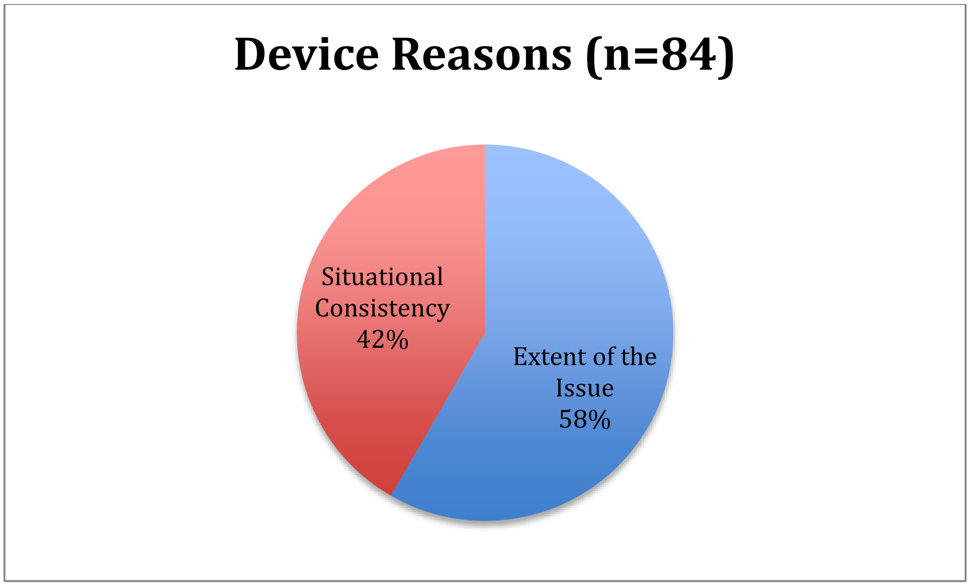


Figure 13. Device Reasons participants provided for carrying out a response in the TSI.

To delve deeper into understanding how users respond to automation issues, we examined which categories of reasons appeared most frequently for which categories of responses (Table 16). Although cell size was too small to test categories of strategies with categories of reasons (even at the higher level of Person, Situation, and Device Reasons, and even if excluding the Wait for Something to Happen strategy group), we did conduct a χ^2 test on the total number of reasons to engage in each strategy group. The numbers of expected cues was calculated to be the same percentage of action sets. For example, 19% of the action sets were of the Continue to Use strategy group and therefore 19% of the reasons were expected to apply to Continue to Use strategies. The χ^2 test revealed that the frequency of reasons did not differ from expectations, χ^2 (5, n = 647) =9.09, p = .106. Therefore, it is not surprising that the most frequently reported strategy group, Change Usage Pattern, was also the strategy group that had the greatest number of reasons (n=200). More than half of these reasons (n=127) were Person Reasons. For example, one participant decided not to rely on the automation's tracking of his activity and instead suggested to "*enter in the information yourself*" because of the Person

Table 16

Reasons for Actions in Strategy Groups in the TSI

	Continue to Use (n[1]=54)	Change Usage Pattern (n=95)	Gather Information or Seek Help to Get it Fixed (n=41)	Wait for Something to Happen (n=5)	Change or Monitor My Behavior in the Situation (n=54)	Try to Fix it on My Own (n=28)	Totals of each Barrier Reason (n=277)
			Person Reasons				
Figuring Out What Caused the Issue	11	21	10	6	44	16	108
Non-causal Knowledge/Experience	10	17	6	1	19	10	63
General Life Knowledge/Experience	13	19	18	0	19	16	85
Importance of Issue/Purpose of Use	42	54	7	3	16	5	127
Personal Preference	5	12	5	1	13	6	42
Awareness	3	4	1	0	3	0	11
Person Reasons Total	84	127	47	11	114	53	436
			Situation Reasons				
No Alternatives	9	12	6	3	2	5	37
Ease of Implementation	22	25	2	2	11	8	70
Cost	0	9	3	0	4	2	18
Total Barrier Reasons	31	46	13	5	17	15	127

Table 16 (continued)

	Continue to Use (n[1]=54)	Change Usage Pattern (n=95)	Gather Information or Seek Help to Get it Fixed (n=41)	Wait for Something to Happen (n=5)	Change or Monitor My Behavior in the Situation (n=54)	Try to Fix it on My Own (n=28)	Totals of each Barrier Reason (n=277)
			Device Reasons				
Situational Consistency	2	7	7	0	6	3	25
Extent of the Issue	14	20	17	1	7	0	59
Total Device Reason	16	27	24	1	13	3	84
			All Reasons				
Total of Reasons in Each Strategy Group	131	200	84	17	144	71	647

[1] n is the number of action sets within each type of strategy group

Reasons of that included knowledge about the activity tracker (e.g., "*because it lets you*") and because it met with his purpose of use "*and because that way I can remember if I actually did work out that day or not.*" A different participant explained that he would change his usage pattern of using the activity tracker with a backup of a secondary tracking app *"I also use MyFitnessPal. That will track my food*" because of the Situation Reasons of "*because you can put in specific foods and recipes… I can put in anything I eat or drink, and it's on my phone so it's very convenient*" and "*it's free*." Device Reasons were also sometimes given for Change Usage Patterns for strategies, such as one participant who suggested writing down calories and activities on paper because "*there's no way it's going to be deleted.*"

The strategy group with the second greatest number of reasons was Change or Monitor My Behavior in the Situation (n=144). Within that strategy, most of the reasons were Person Reasons (n=114). For instance, one participant described his experience with the issue of the battery dying without accurate warning. He recommended changing his behavior to try to "*just charge it every day*" because of his knowledge about the typical battery life of his activity tracker (i.e., regardless of whether the tracker gives warning or not) that the "*battery dies in about 7 days, I guess, roughly. I charge every day or every other day, it should be fine because it's still within that battery-dying time."* Situation Reasons were less common for the Change or Monitor My Behavior in the Situation (n=46), but they did still emerge. For instance, one participant described a scenario in which he moved his arm while remaining on the same step of an escalator and observed that his step count increased. He suggested that the issue could be avoided by changing his behavior back to *"riding the escalator like a normal person*" (i.e., not swinging his arm). For instance, if he was in a situation when he could not climb the stairs/escalators, "*if I had the luggage in tow*" or "*if there were people standing kind of blocking the escalator*" he would not swing his arm and not climb the escalator. Device Reasons were also mentioned on occasion for Change or Monitor My Behavior in the Situation. For instance, one participant said he might uninstall several of the apps on his phone if the synching issue was consistent over time.

Many reasons were also provided for Continue to Use strategies. Again, Person Reasons were the most frequent (n=84) and included facets of knowledge, Affect, and Purpose of Use/Importance of the Issue. For instance, one participant described how the automation issue did not really impact her purpose of use, "*because I use this tool in general just to get ballparks, I'm not really rigid about the fact that it counted 50 extra steps a day, you know, I realize the limitations, so it's okay if it's not perfect*" and so she would respond by saying *"'ehh, that's the way it is'...and just kind of moved on."* Another participant described using Situation Reasons for not "*do anything about it {the issue}, just kind of like 'Okay, that's strange.'*" The Situation Reasons suggest the action was "*the easier thing*" to do and then he did not "*have to take the time out of the day*." Again, Device Reasons were rare. As an example of a Device Reason, one participant said he would *"keep using* it" because the extent of the error was not that substantial, "*it's not a huge number of steps* {that the tracker is off by}."

Fairly similar Person, Situation, and Device Reasons emerged for the three smaller strategy groups of Gather Information or Seek Help to Get it Fixed, Try to Fix it on My Own, and Wait for Something to Happen. The only obvious difference was that for the Gather Information or Seek Help to Get it Fixed, the typical pattern for strategy groups, wherein most of the reasons were Person Reasons, followed by Situation Reasons, followed by Device Reasons, did not hold. Rather, there were more than double the number of Device Reasons than Situation Reasons. A more detailed look revealed that this was because participants frequently wanted to check for situational consistency when gathering information and seeking help, such as to see if other users reported online having the same issue in the same situations. For instance, one participant said he would google search the issues "*and see if there's someone with similar issues. Usually there's always someone with similar issues."* In sum, most types of strategies used a variety of reasons, with Person Reasons playing a particularly prominent role in strategy selection, but responses of Gather Information or Seek Help to Get it Fixed seemed to especially take into account Situational Reasons as well.

Summary of TSI Results

In sum, participants generated many incidents that could make it difficult for them to use their activity trackers. Many, but not the majority (i.e., >50%), of their explanations about the nature of those incidents were Causal Automation Error interpretations. The issue interpretation details were related to the cues to issues. Participants most frequently name responses related to changing their usage patterns, to continuing to use the activity tracker, and to changing or monitoring their own behavior in in the situation. Person Reasons were the most frequently mentioned reasons to carry out, and to not carry out, an action. Qualitatively, reasons for responses were related to responses. In short, the TSI provided a glimpse in the automation issue troubleshooting process.

CHAPTER 5

SCENARIO-BASED INTERVIEW RESULTS

The results of the Scenario-Based Interview (SBI) were examined to answered four research questions related to troubleshooting attended automation issues:

- [R2] With what level of detail do experienced everyday automation users interpret automation issues?
- [R3] Upon attending to an automation issue, how do experienced everyday automation users decide to respond?
- [R4] What strategies do experienced everyday automation users have for responding to an automation issue?
- [RS] To what extent do users' device mental models relate to how they interpret automation issues?

Participants were given five separate scenarios that contained signs of an automation issue. For each scenario, participants rated the severity (impact) of the issue and their confidence that the scenario was an automation issue. Participants also described the nature of what they thought had occurred in the scenario and what they would do in the scenario. Data segmented and analyzed included issue interpretation details, reasons for response selections, and responses to automation issues. Segmentation definitions and coding rules were the same as those of the TSI, except for where otherwise noted.

Scenario Descriptions

Responses to "What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker?" were recorded for each participant for each scenario. Responses were on a 4-point scale of (1) no impact, (2) minor impact, (3) moderate impact, and (4) serious impact. Table 17 reports the descriptive results of the severity of each scenario. In general, scenarios were rated as having between a minor and moderate impact. The only exception to this was Scenario E, a situation in which the automation is completely unresponsive. As expected, this situation was rated almost always as having a serious impact on the user's ability to use his or her activity tracker.

Table 17
Impact of the Scenario on Ability to use the Activity Tracker

	Scenario A	**Scenario B**	**Scenario C**	**Scenario D**	**Scenario E**	**Across Scenarios**
M	3.13	2.13	2.23	2.7	3.93	2.8
SD	0.97	0.68	0.9	0.75	0.25	1.0
Min	1	1	1	1	3	1
Max	4	4	4	4	4	4

Responses to "How confident are you in your judgment of this situation (that is whether or not this is an issue)?" were also recorded for each participant for each scenario. Responses were on a 6-point scale of (1) extremely confident it is not an issue, (2) moderately confident it is not an issue, (3) slightly confident it is not an issue, (4) slightly confident it is an issue, (5) moderately confident it is an issue, and (6) extremely confident it is an issue. Table 18 reports the descriptive results of issue confidence for each scenario. In general, participants reported being confident (slightly, moderately, or extremely) that the scenario described an issue (i.e., instead of the scenario not being an issue). Participants were particularly confident that Scenario E represented an automation issue.

Table 18
Confidence that the Scenario Describes an Issue

	Scenario A	**Scenario B**	**Scenario C**	**Scenario D**	**Scenario E**	**Across Scenarios**
M	4.13	3.43	3.97	3.97	5.77	4.3
SD	1.57	1.52	1.59	1.52	0.77	1.6
Min	1	1	1	1	3	1
Max	6	5	6	6	6	6

Issue Interpretation: [R2] With What Level of Detail Do Experienced Everyday Automation Users Interpret Automation Issues?

To examine [R2] the transcripts to the following interview questions were segmented:

1. Can you please describe what you think is going on in this scenario?
2. What do you think the issue is?
   a. *If error that is specific or has a cause*, what would suggest that was the root of this issue?

   OR

2. Why do you not think this indicates an issue with the {activity tracker}?

The same multi-level coding scheme utilized in the TSI was utilized for the SBI:

- Level 1: Not an Automation Error or Automation Error
- Level 2: General Automation Error or Specific Automation Error
- Level 3: Non-Causal Automation Error or Causal Automation Error

Additionally, the 6-point confidence scale was also occasionally utilized to help differentiate between Not an Automation Error and Causal Automation Error. The scale was utilized most frequently when it was uncertain if the participant was describing a device limitation (e.g., unable to accurately sense a particular activity because of a programming or sensing limitation) or an automation issue with a cause. To elaborate, the device limitation could be considered Not an Automation Error if the participant thought the device should not be responsible for tracking the activity. However, the issue was considered an Automation Error if the participant understood the limitation as the cause for the automation error. Thus, only in cases when the researcher could not determine which was the most appropriate code to apply, the researcher consulted the confidence scale. Confidence ratings of 1, 2, or 3 allowed the researcher to classify the issue detail as Not an Error and ratings of 4, 5, and 6 allowed the researcher to classify the issue detail as a Causal Automation Error. Two independent-raters reached inter-rater agreement greater than 80% both across scenarios and participants before splitting up the remaining transcripts.

From the 5 scenarios and 30 participants, 273 issue details were segmented and coded. The mean number for issue details per scenario was 54.60 (SD=10.69; Table 19). Scenario E, which

Table 19
Number of Issue Details, Actions, and Reasons per Scenario

Scenario	**Issue Details**	**Actions**	**Reasons**
A	45	66	102
B	51	46	72
C	52	59	90
D	52	54	74
E	73	70	112

was also intended to be the least concrete scenario, had the greatest number of issue details at 73. Most (80%) of the interpretation details generated by participants were automation errors. A chi-square goodness of fit test demonstrated that Automation Errors details were overrepresented compared to Not an Automation Error, χ^2 (1, N = 273) = 97.32, $p < .001$. This pattern held for every scenario (Table 20; Figure 14). As a whole, the scenarios followed the pattern of having more Specific Automation Error issue details then General Automation Error details, χ^2 (3, N = 218) = 37.16, $p < .001$). However, this pattern did not hold true for Scenario E, the most vague

Table 20
Issue Interpretation Details Across Scenarios

	Scenario A	Scenario B	Scenario C	Scenario D	Scenario E	Total
Frequency						
Not an Automation Error	10	14	9	16	6	55
Automation Error	35	37	43	36	67	218
General Automation Error	6	5	9	8	36	64
Specific Automation Error	29	32	34	28	31	154
Non-causal Specific Automation Error	13	6	14	13	12	58
Causal Specific Automation Error	16	26	20	15	19	96
Test						
$\chi 2$ Not an Automation Error vs Automation Error	$p<.001$	$p=.001$	$p<.001$	$p<.006$	$p<.001$	$p<.001$
$\chi 2$ General Automation Error vs Specific Automation Error	$p<.001$	$p<.001$	$p<.001$	$p=.001$	$p=.541$	$p<.001$
$\chi 2$ Non-causal Specific Automation Error vs Causal Specific Automation Error	$p=.557$	$p<.001$	$p=.303$	$p=.705$	$p=.209$	$p=.002$
$\chi 2$ Not an Automation Error, General Automation Error, Non-causal Specific Automation Error, Causal Specific Automation Error	$p=.182$	$p<.001$	$p=.098$	$p=.404$	$p<.001$	$p=.001$

scenario, where the number of Specific Automation Error details was not significantly different than the number of General Automation Error details. Within only Specific Automation Error details, when considered total across all scenarios, there were significantly more Causal Specific Error details then there were Non-causal Specific Error details. However, this was not typically obvious at the scenario level, where the difference was only significant in Scenario B.

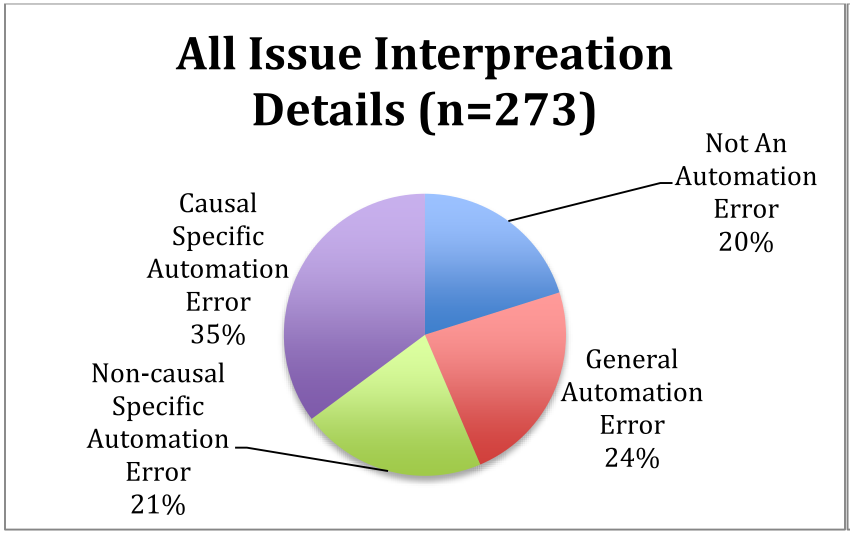


Figure 14a. All issue details across all scenarios

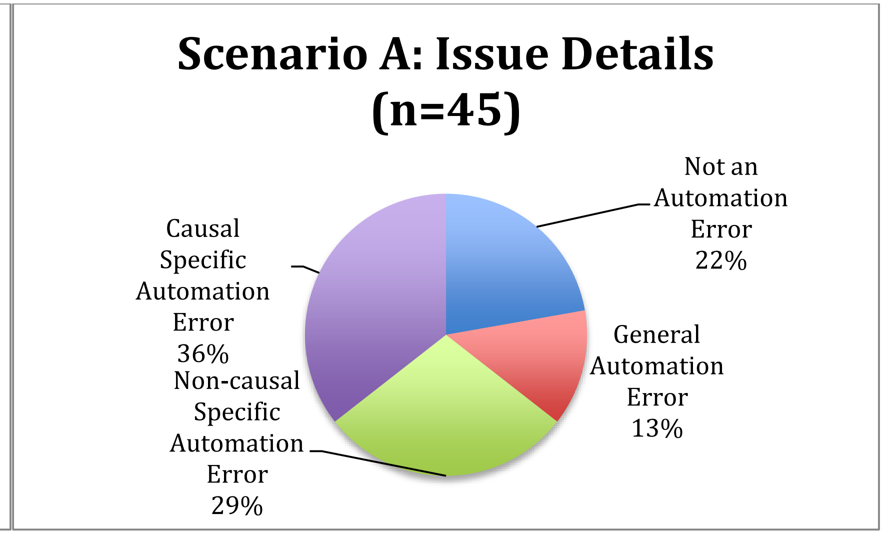


Figure 14b. Issue details in Scenario A

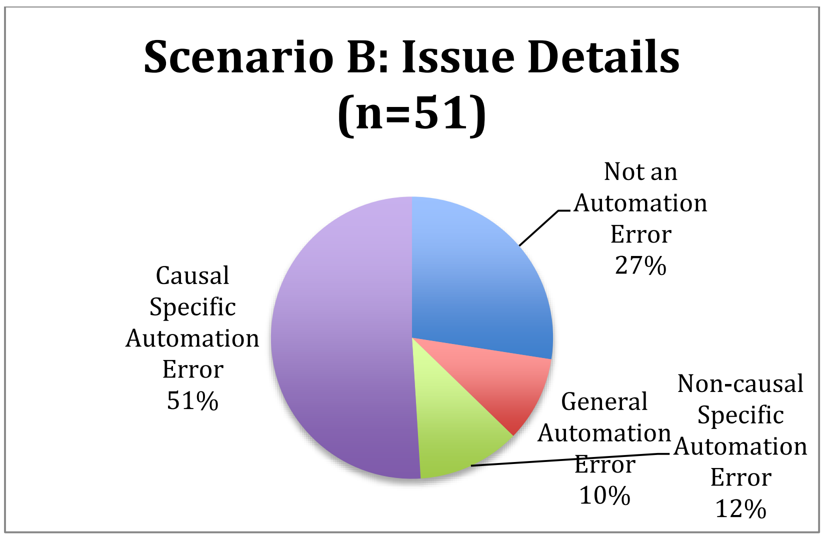


Figure 14c. Issue details in Scenario B

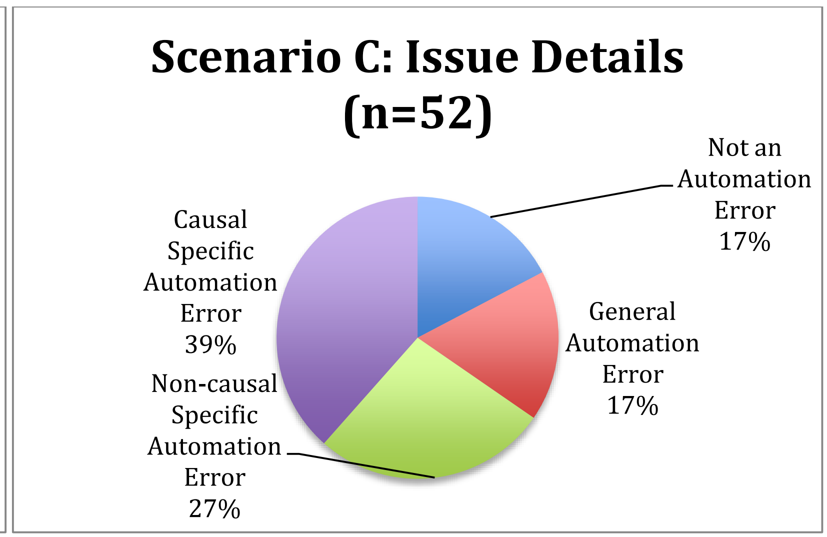
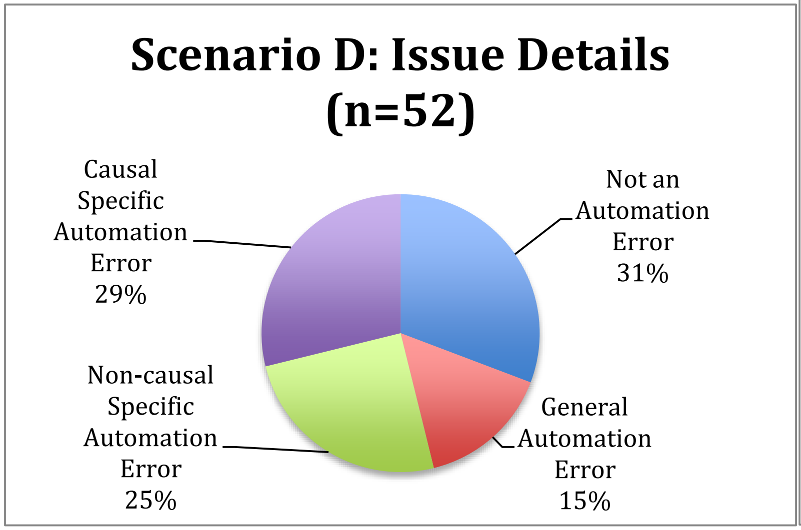


Figure 14d. Issue details in Scenario C

Figure 14. Issue interpretation details in the SBI scenarios.



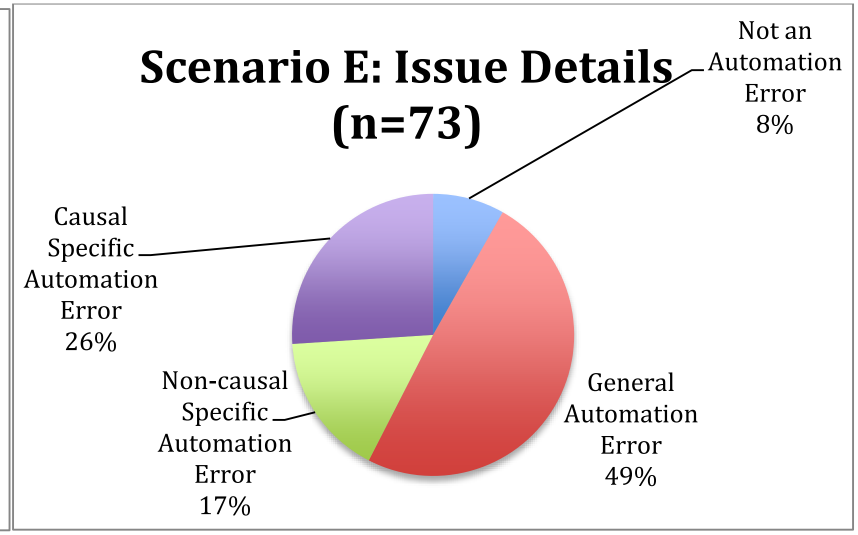


Figure 14e. Issue details in Scenario D

Figure 14f. Issue details in Scenario E

Figure 14 Continued. Issue interpretation details in the SBI scenarios.

Figure 14 shows the breakdown of error types. As a whole, Causal Specific Errors were the most frequently reported issue interpretation (35%). However, other interpretations were prevalent and each made up 20% or more of issue interpretation details. A χ^2 test for independence revealed that scenario and issue detail type were not independent, $\chi^2(12, N = 273) = 50.28$, $p < .001$. To help provide a more concrete comparison of the four types of issue interpretations participants made, Table 21 provides quotes of each detail code from Scenario D.

Descriptively, scenarios B and D were more frequently classified as Not an Automation Error then were A, C, and E. In looking at the data, users in B and D often thought the user had made an error. For instance, one participant in Scenario B explained that the automation's distance reading might differ from that of the known trail because "*maybe I even made it a little bit longer than it actually*

Table 21

Issue Interpretation Detail's Coding Scheme From Scenario D: Imagine you are reviewing your sleep data and find that your activity tracker says you only awoke once throughout the night. However, you recall waking up several times.

Issue Detail Code			**Definition**	**Exemplar Quote**
Not a Automation Error			The participant explains that the device is supposed to work that way and thus it is not an error or the participant explains that the user made a mistake**.**	"If you're not getting up and actively moving around, it's probably just going to think it's minor movement that you do in your sleep anyway, so chances are, it's kind of within operating conditions for sleep mode."
Automation Error	General Error		The participant states that an error, problem, issue, or other general term occurred. The participant does not articulate what exactly the error is or where it occurred. This might include, but are not limited to, situations where the participant explains that the error is ongoing or the error somehow makes the tracker inaccurate or reliable (e.g. over-estimation, under-estimation).	"I think it's malfunctioning."
	Specific Error	Non-Causal Error	The participant explains *where* or *what specific* feature or function erred. These are not limited to, but might include the distance, step counter, stair counter, sleep tracker, progress chart, calories, synching, or other specific feature or function. If what, the participant could explain that the issue ranges across all functions (e.g., a calibration issue).	"I guess it's another flaw in its heart rate sensor or detector."
		Causal Error	The participant explains how the error occurred or what caused the error. These may include a variety of specific errors with a logical explanation, information about how	"I know it senses when it is being moved more than the certain area in a certain amount of time. So if you

			the device functions, damage to the entire tracker, and many other explanations. It typically includes a *how or why*, not just a *where* or *what.*	wake up and you don't move the arm that you're having your tracker on, it likely would not sense that."

was and didn't follow it exactly." Likewise, a participant explained that the sleep tracker's automation might not match the number of times she recalled waking up in Scenario D because "*it could honestly be something with me personally. Where, umm, you, your body isn't in REM sleep, or deep sleep, but it's not fully awake either. And so, there have been times where I thought I was awake much more than I was. And in reality, I think I was still slightly asleep, but not quite awake."* In contrast, such explanations were far less likely in situations where the participant was described as fully awake and alert (i.e., exercising), situations where the issue happened repeatedly, and in situations of complete device malfunction.

General Automation Error Interpretations were especially prevalent in Scenario E compared to the other scenarios. Scenario E, which described a completely unresponsive device, was indeed designed to be vague. Participants tended to interpret the error generally because they were unsure what was going on. For instance, one participant simply explained, "*I guess I don't know what the issue is, but its just died. It just refused to do anything*" and another participant echoed this interpretation, "*it's broken.*"

In looking at Non-causal Specific Automation Error details, Scenarios A and C seemed to have proportionally more details than Scenarios B, D, and E. However, these differences did not seem drastic, as all 5 Scenarios were less than 10% different in their Non-causal Specific Automation Error details from the total number of Non-Specific Automation Error details. These interpretation details tended to explain what parts of the activity tracker were impacted by the scenario, but not why. For instance, in Scenario C wherein the automation's distance repeatedly differed from a known distance, participants explained "*it's the accelerometer not being accurate*" but not why the accelerometer was not accurately counting a step (e.g., running up hill vs down hill, etc). Similarly, in Scenario A wherein calories did not increased after a vigorous work-out, one participant explained, "*I would say the issue would be either it's malfunctioning* *and how it reads your heart rate and then calculates calories burned…"* but did not explain why the heart rate was malfunctioning (e.g., poor contact with the skin, reflection of the light due to sweat).

Descriptively, the more concrete scenarios of A, B, and C had more Causal-Specific interpretation details then did D and E. For instance, in Scenario A, participants frequently explained that the calorie count was not increasing because no motion was detected because the tracker was worn on an arm. Because the arm was not moving, neither was the activity tracker, *"especially in spin, your arms generally don't move at all cause your using your arms as stability, so it's not going to necessarily sense any movement, cause all of the movement and energy you're expending is going to be in your legs. So it's just fairly simply not doing anything."* In Scenario B and C, Causal-Specific interpretations frequently explained why the stride length was not calculated correctly, as this participant explained in Scenario B, "*That would probably be using the step counting function, and you might have a shorter or longer stride than the average human, so it's going to count more or less steps, in this case steps, or distance covered per step, than you actually did."* In Scenario B and C, participants also frequently made the Causal-Specific interpretation that the activity tracker might be adding steps and distance because of too much sensitivity or increased hand movement, as these two participants explained in Scenario C, "*maybe you have it on your dominant hand, because it changes the sensitivity depending on what hand you have your tracker on so that might be user error*" and this participant *"I think again, that it's picking up extra arm movements. That it's just throwing it off, it's over counting."*

In sum, all scenarios lead to interpretations of Not an Automation Error, General Automation Error, Non-causal Specific Automation Error, and Causal Specific Automation Error. Causal Specific Automation Error was the most frequent interpretation overall, but this finding varied by scenario. For very vague but high-impact automation issues, such as Scenario E, General Automation Error interpretations were more frequent. Clearly the scenario is related to how users interpret automation issues.

Strategies for Responding to Automation Issues and Reasons for Strategy Selection

To examine [R3] and [R4] the transcripts to the following interview questions were segmented:

1. How would you react to this situation to be able to track your activity, fitness, or health?

   <continue if the response does not include a why>

   i. Prompt: What would you do and why?

   ii. Prompt: what would make want to try that strategy for coping with this situation?

2. *If applicable*, Can you tell me what would you do if that did not work and the issue kept happening?

   <if the response does not include a why>

Prompt: what would make want to try that/those/these strategy for coping with this situation?

The segmenting rules for a strategy action 'set' were based on the TSI results. In particular, actions were segmented in the same way, and excluded redundant actions. Because only 4 action sets in the TSI contained more than one action (and 273 contained only one action), an action sort (i.e., for each incident for each participant) was not conducted and "action sets" were considered the same as an "action." A coding scheme was developed from the TSI action set sort, which resulted in 37 strategies. These 37 strategies were defined by the name given to the action groups (e.g., "manually keep track on paper or excel") and examples from the TSI. These 37 strategies were divided into the same 6 strategy groups as the TSI. Additionally, a 38th strategy of "other" was added for coding the SBI. The complete strategy coding scheme is available in Appendix N. The operational definition of a “reason” and the coding scheme for reasons were the same as the TSI. However, no general reasons (i.e., a reason without reference to a specific action or set of actions) and only 6 barrier reasons were mentioned in the SBI. Therefore, only reasons to carry out an action are presented here. Two independent-raters reached inter-rater agreement greater than 80%, both across scenarios and participants, on both reasons and strategy coding schemes, before coding the remaining transcripts.

Strategies for Responding to an Automation Issue: [R4] What Strategies do Experienced Everyday Automation Users Have for Responding to an Automation Issue?

Twenty-seven (of the 38) strategies emerged in the SBI in a total of 295 actions. Participants reported between 5 and 17 actions each, a mean of 9.83 per participant (and mean 1.97 per participant per scenario). The mode of actions per participant was 8 (SD= 2.57). Combining across participants and looking at each scenario, the mean number of actions per scenario was 59 (Table 19). Two actions were classified as “other” and were thus eliminated from the rest of data analysis. Both of these eliminated actions described writing a bad review of the product. Additionally, only one action fell into the Wait for Something to Happen strategy group. Because this group was so small, it was also eliminated from further data analysis. In examining the 5 remaining strategy groups across scenarios, a chi-square goodness of fit test revealed that Change Usage Pattern, Gather Information or Seek Help to Get it Fixed, and Continue to Use strategies were over-represented and Change or Monitor my Behavior in the Situation and Try to Fit it on My Own were underrepresented, χ_2 (4, N = 292) = 69.64, $p < .001$. Almost every scenario contained one action of each of the 5 remaining strategy groups (Table 22). The only exception to this was that the strategy Continue to Use was never mentioned in the most severe scenario, Scenario E. Because of cell size, additional chi-tests at the scenario and item level were

Table 22
Number of Actions in Each Scenario

Row Labels	**Scenario A**	**Scenario B**	**Scenario C**	**Scenario D**	**Scenario E**	**Total number of each action code**
Continue to Use (n=76)	11	24	22	19	0	76
Change Usage Pattern (n=86)	26	6	8	13	33	86
Gather Information or Seek Help to Get it Fixed (n=82)	14	8	19	11	30	82
Change or Monitor My Behavior in the Situation (n=30)	12	2	5	9	2	30
Try to Fix it on My Own (n=18)	2	6	5	1	4	18
Total actions per scenario	66	46	59	53	69	293

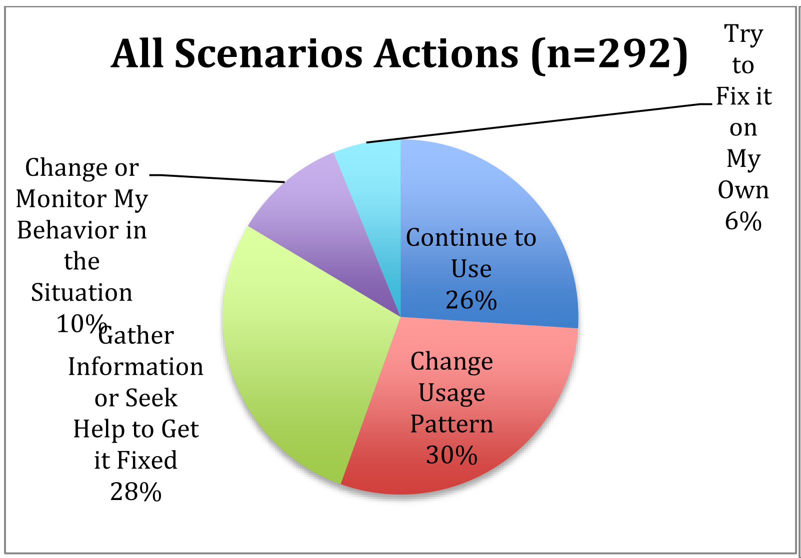


Figure 15a. Actions in all scenarios

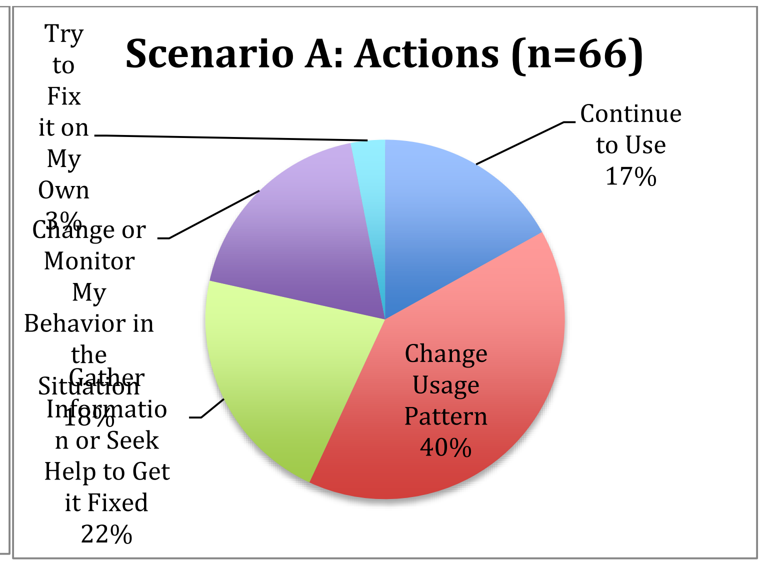


Figure 15b. Actions in Scenario A

Figure 15. Strategy groups in the SBI across scenarios

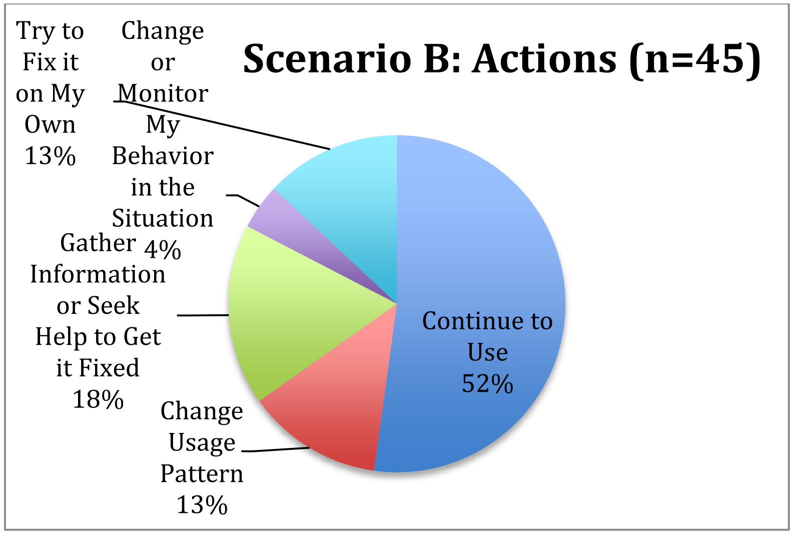


Figure 15c. Actions in Scenario B

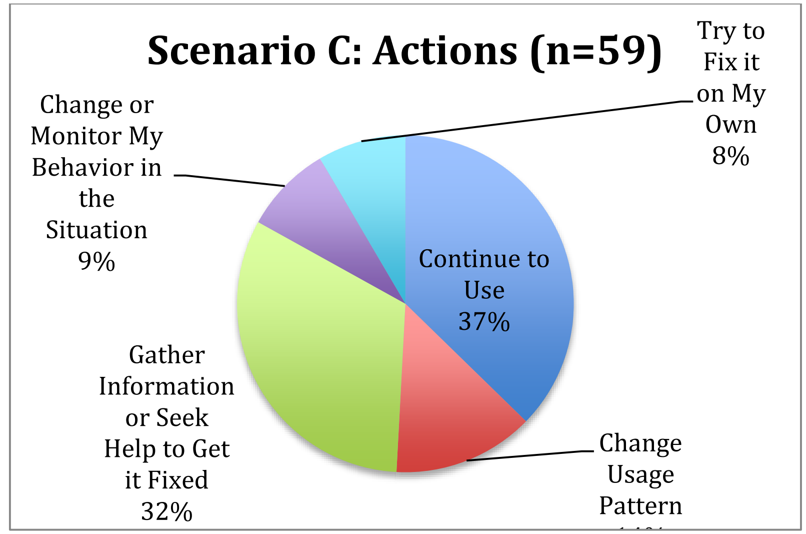


Figure 15d. Actions in Scenario C

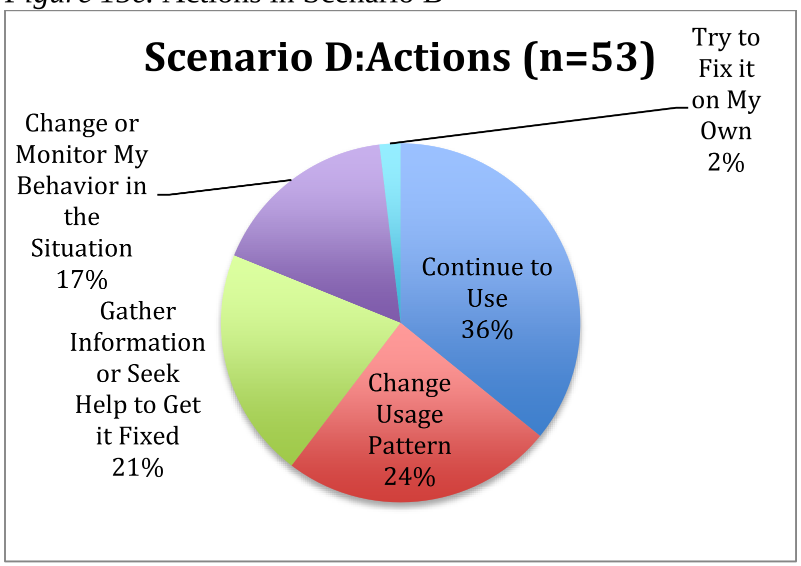


Figure 15e. Actions in Scenario D

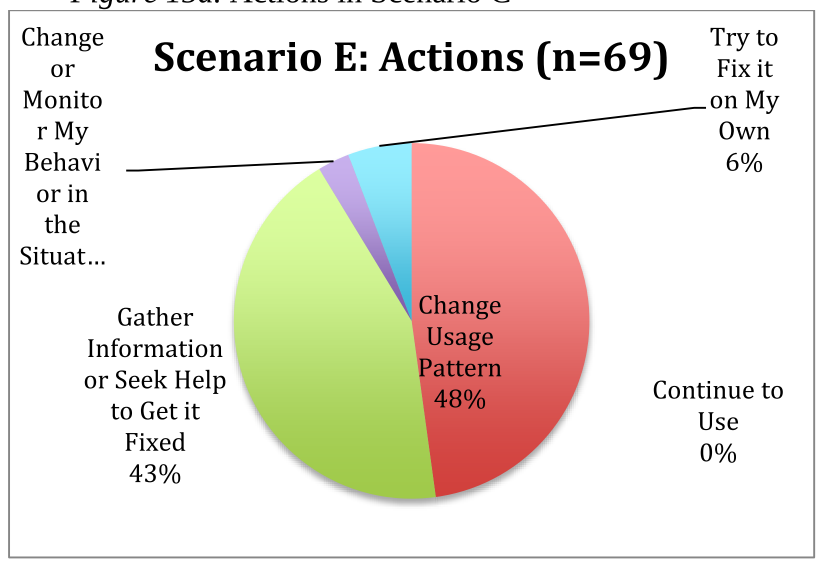


Figure 15f. Actions in Scenario E

Figure 15 (continued). Strategy groups in the SBI across scenarios not conducted for scenarios. Rather, the reader should reference Table 22 and Figure 15 for descriptive patterns.

The most frequently mentioned strategy group across all scenarios, Change Usage Pattern, was only the most frequently mentioned strategy in two of the scenarios: Scenario A and Scenario E. Scenario A described a substantial issue (3.13 out of 4 on the impact scale) on a feature used by many participants (i.e., compared to Scenario D's sleep tracker, which not as many participants reported using). Additionally, Scenario A provided a context where other alternative tracking options (e.g., exercise machine) were readily available. Indeed, many participants described not relying on their activity tracker and rather recommended using a different technology and manually adding in the exercise as these participants explained, "*Well I guess sometimes if you do those machines in the gym, they sometimes also give you an idea of how many calories you burned from the machines and you can get an idea from there*" or *"manually enter into my activity log that I did spin class for an hour."* Scenario E described a non-responsive device, wherein many participants reported they would no longer be able to use their activity trackers, "*I would probably have to find a different tracker or replace this one."*

In looking at the second greatest strategy group, Gather Information or Seek Help to Get it Fixed, was also the second most frequently mentioned type of strategy for every scenario. However, Gather Information or Seek Help to Get it Fixed seemed to play a particularly obvious role in Scenario E, where it made up 43% of responses. One participant described his response "*I would go on like the customer service*" and another participant said he *"would probably email the company.*" Again, Scenario E was the most severe and most vague scenario, so it may have been related to participants not knowing what was the issue.

The third most frequently mentioned strategy group was to Continue to Use the activity tracker. This strategy group was particularly frequent in Scenarios B and C, where Change Usage

Pattern was not a frequent strategy group. Participants described correcting for the discrepant distance reading in their mind, such as this participant did in Scenario B, "*I think I would take in the information, but then I will say to myself it's an average it's not exact. And I would just have an idea of what my activity usage is.*" Participants in Scenario C echoed the strategies of continuing to use the activity tracker, "*I'd most likely still ignore it.*" The Continue to Use strategy group also emerged fairly frequently in Scenario D, where one participant described her response as, *"I would let it go. Just leave it is at is."* Lastly the scenarios with frequent mention of Continue to Use strategies, Scenarios B, C, and D, were the same scenarios that were reported as having the least amount of impact on the user's ability to track their activity.

The second least frequently reported strategy group was Change or Monitor My Behavior in the Situation. This strategy group was particularly prominent in Scenarios A and D compared to Scenarios B, C, and E. Both Scenarios A and D described a sensitivity issue wherein the activity tracker may not correctly detect movement or stillness. Participants frequently described making adjustments so that the tracker could more accurately detect motion. For example, this participant in Scenario A said if, "*I went to spin class, I would make sure it's in the correct position.*" In Scenario D, participants also discussed monitoring their own behavior, perhaps to explain the discrepancy between what the activity tracker stated and what they thought had happened while they were asleep, "*I would probably try to pay more attention to what I'm doing when I wake up in the middle of the night before I consider it to be a fault of the device.*"

The least frequently reported strategy group was Try to Fix it on My Own. Although it is difficult to conclude discernable patterns with such a small group of actions (n=18), Try to Fix it on My Own did appear to be more frequent in Scenario B than in other scenarios. Participants in Scenario B explained that they thought a setting of stride length was incorrect and could be fixed by user input. For instance, this participant explained, "*I could change stuff for the stride length*" and another participant said "*I would first check and make sure that the app had my correct height in there.*" Nonetheless, Try to Fix it on My Own strategies did appear in other scenarios, and tended to include changing setting inputs or restarting or resetting the activity tracker.

Reasons for Responses: [R3] Upon Attending to an Automation Issue, How do Experienced Everyday Automation Users Decide to Respond?

Four hundred and fifty reasons to carry out an action were segmented and coded from the SBI. The number of reasons reported by participants ranged from 5 to 28, with 13 being the most frequently reported number of reasons per participant. A mean of 15.00 reasons (SD=5.89) were reported per participant (i.e., a mean of 3.00 reasons per scenario per participant or a mean of 1.53 reasons per action). Each scenario also had a mean of 90.00 reasons (Table 19). Five reasons did not fit into the reasons coding scheme and were therefore eliminated from the reminder of data analysis.

Across scenarios, Person Reasons were the most frequently reported reason type compared to Situation and Device Reasons, χ^2 (2, N = 445) = 183.40, $p < .001$(Figure 16). Situation Reasons (n=82) were mentioned about as often as Device Reasons (n=80). Additionally, the distribution of Person, Situation, and Device reasons were dependent on the scenario, χ^2 (8, N = 445) = 187.80, $p < .001$ (Table 23). The chi-square test of independence revealed that Scenario A had more Person Reason and less Situation and Device Reasons than expected and that Scenario B had an overrepresentation of Situation Reasons and underrepresentation of Person and Device Reasons. Scenario C had an overrepresentation of Device Reasons and an underrepresentation of Person and Situation Reasons. Scenario D had an overrepresentation of Person Reasons and an underrepresentation of Situation and Device Reasons. Finally, Scenario E had an overrepresentation of Situation Reasons and an underrepresentation of Person and Device Reasons.

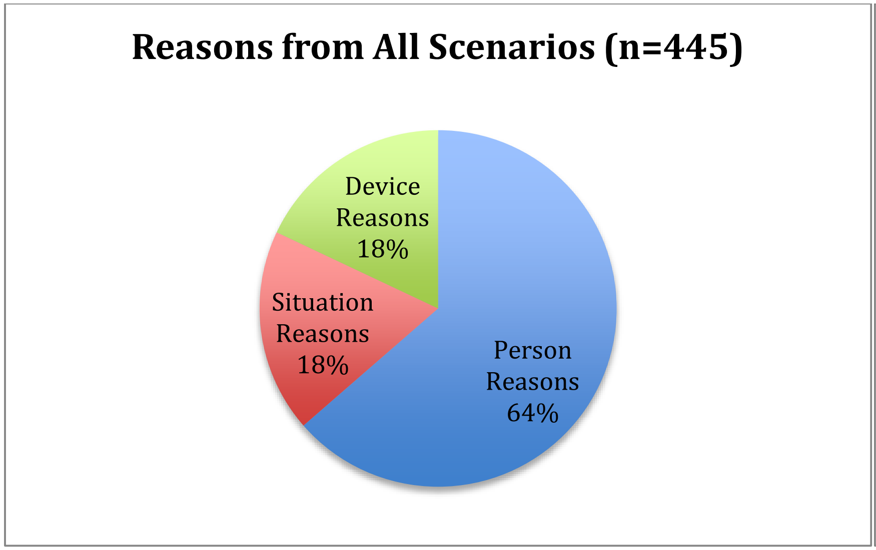


Figure 16a. Reasons for actions across all scenarios

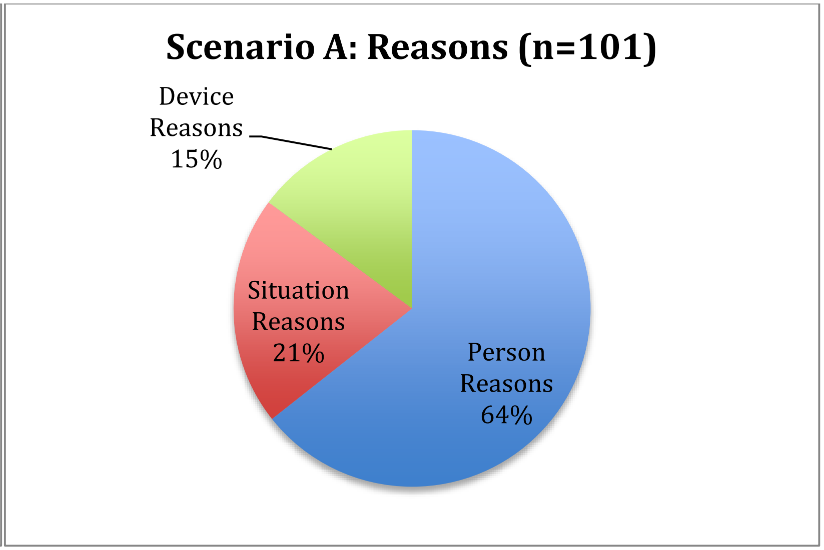


Figure 16b. Reasons for actions in Scenario A

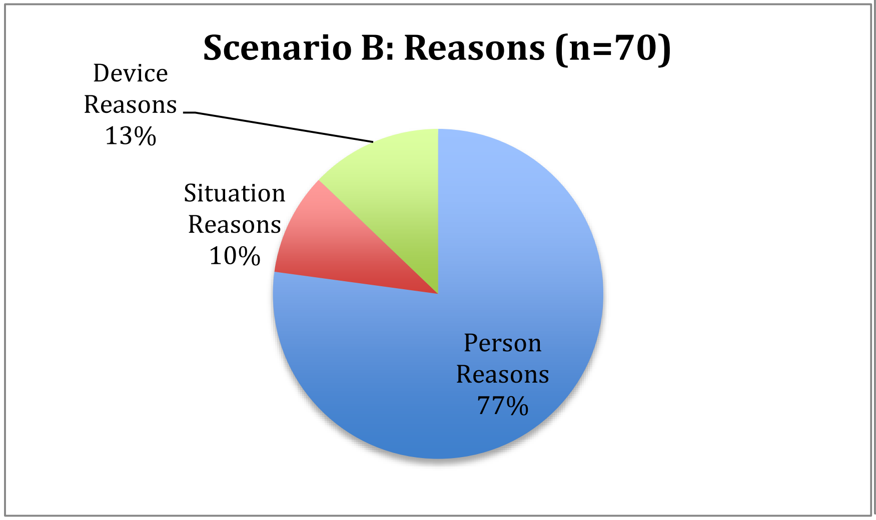


Figure 16c. Reasons for actions in Scenario B

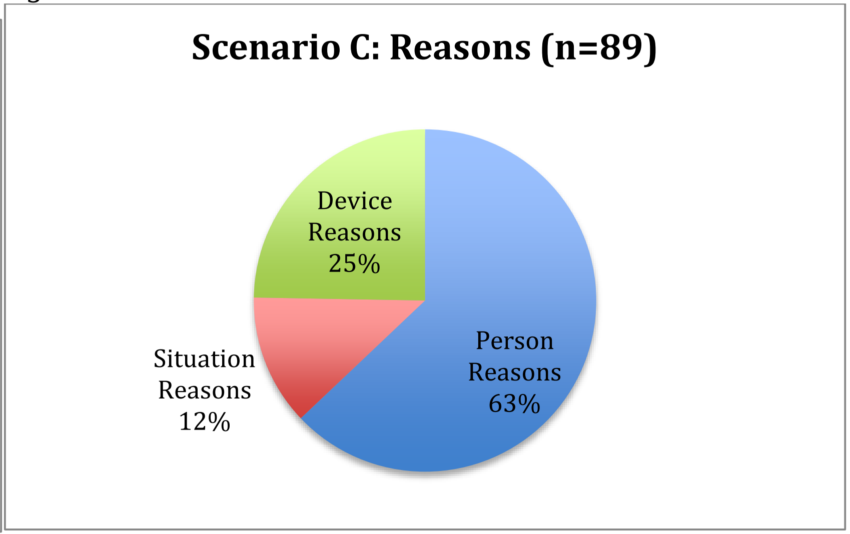


Figure 16d. Reasons for actions in Scenario C

Figure 16. Reasons given for actions in the SBI.

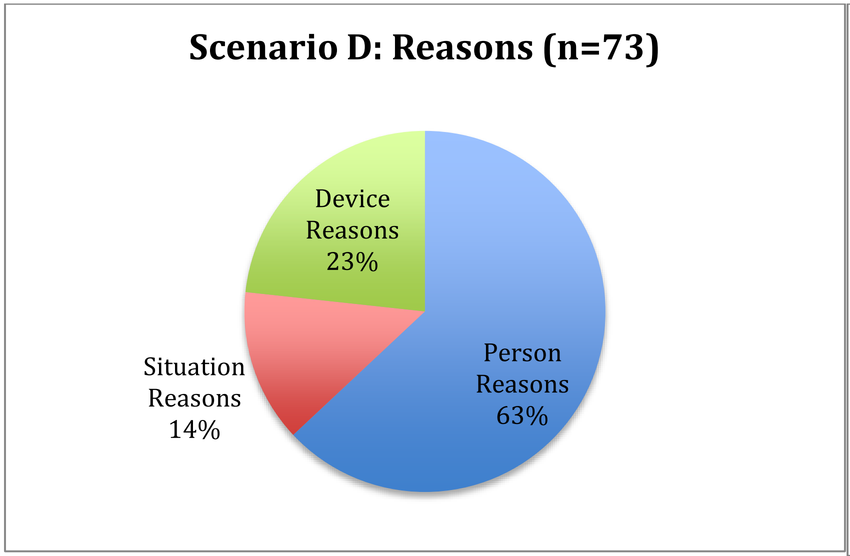


Figure 16e. Reasons for actions in Scenario D

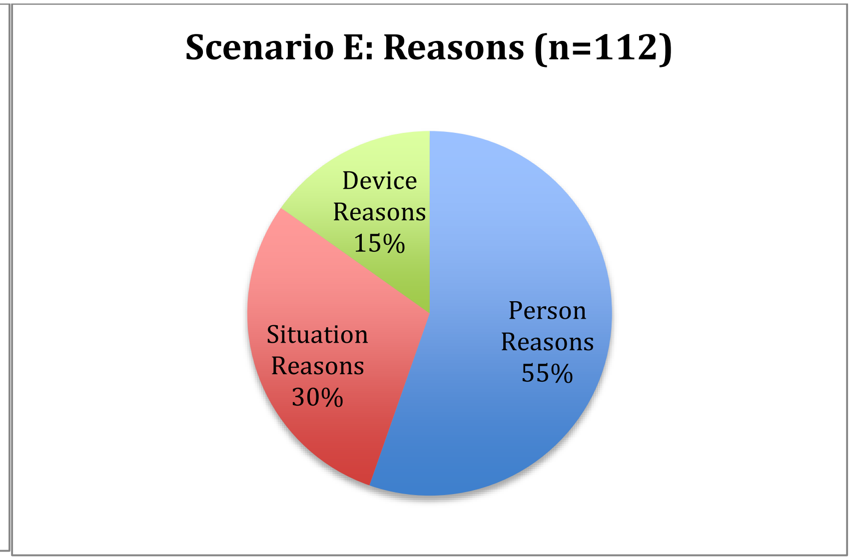


Figure 16f. Reasons for actions in Scenario E

Figure 16 (continued). Reasons given for actions in the SBI.

Person Reasons (Figure 17) included a large variety of codes. In particular, the Importance of the Error/Purpose of Use and Knowledge of various types were frequently used Person Reason Codes. For instance, in Scenario A, one participant explained that she used an activity tracker to track her fitness, and therefore the tracker not picking up her spin class was a substantial problem because "*Because if it's not able to track fitness then there's no point in it being a fitness tracker*." This sentiment was echoed by many participants, but also appeared in reverse wherein the issue was not a substantial problem because of the participant's purpose in using an activity tracker. For instance, in Scenario D, one participant explained his lack of concern with the sleep tracker's issue, "*because sleep tracking isn't anything that I'm super... I don't really care as much about sleep tracking as I do the other activity tracking*." Both

Table 23.
Results of a Chi-square Test for Independence for Scenario and Reasons

		Person Reasons	Situation Reasons	Device Reasons	All Reasons
Scenario A	Observed Frequency	65	15	21	101
	Expected Frequency	39.3	36.1	25.6	101.0
Scenario B	Observed Frequency	7	54	9	70
	Expected Frequency	27.2	25.0	17.8	70.0
Scenario C	Observed Frequency	22	11	56	89
	Expected Frequency	34.6	31.8	22.6	89.0
Scenario D	Observed Frequency	46	17	10	73
	Expected Frequency	28.4	26.1	18.5	73.0
Scenario E	Observed Frequency	33	62	17	112
	Expected Frequency	43.5	40.0	28.4	112.0
All Scenarios	Observed Frequency	173	159	113	445
	Expected Frequency	173.0	159.0	113.0	445.0

General Knowledge and Experience and knowledge related to the error's cause were frequent as well. For instance, a participant in Scenario A might contact customer service because of the life experience with customer service departments where "*they would find the problems and directly solve the problems.*" Causal knowledge reasons or attempts to determine the cause of the issue also included cases were the participant thought the cause of the issue was the user or a device limitation, as this participant in Scenario D explained, "*My initial thought isn't that the device is wrong, just that maybe I'm not moving a whole lot, even though I am awake, like maybe I'll just look in the clock and go back to sleep or something.*"

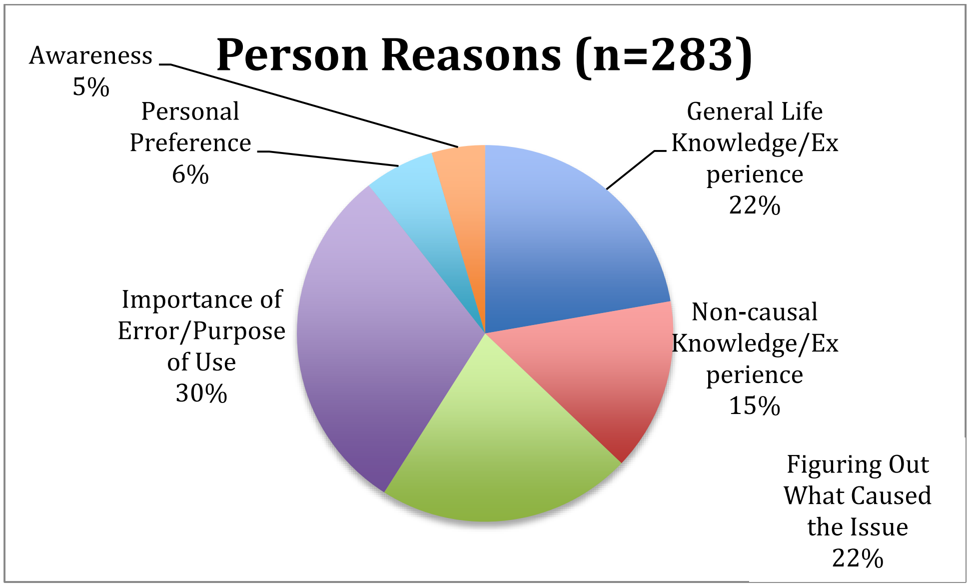


Figure 17. All Person Reasons for actions in the SBI

Situation Reasons (Figure 18) were mostly concerned with the lack of other alternative strategies to try, but ease of strategy implementation and cost also contributed to the decision. An example of a No Alternative reason is, "*Because there's nothing you can do about it"* as why to ignore the issue in Scenario B. Similar No Alternative reasons occurred in Scenario E, "*if it's not responding, there's not much I can do.*" An example of an Ease of Implementation reason from Scenario B is "*because it requires very little work*." Cost reasons typically mentioned the price of the activity tracker, like these participants did in Scenario E, "*I want my $150 tracker to work!*" and "*they are not cheap*."

Device Reasons (Figure 19) were divided into reasons about the Extent of the Issue (e.g., frequency in occurrence, percentage of error deviation from true value) and Situational Consistency (e.g., does the tracker behave the same or different in the same or different situations). Many Device Reasons emerged in Scenario C. For instance, this participant discussed the reasoning of if the issue happened just once or repeatedly, "*I might be a little bit more proactive as far as fixing it if it happens on a regular basis*." As an example of Situational

Consistency in Scenario D, this participant explains trying to determine if the issue occurs more broadly, "*and see if a wider scale, a wider selection, of people had had the same issue*."

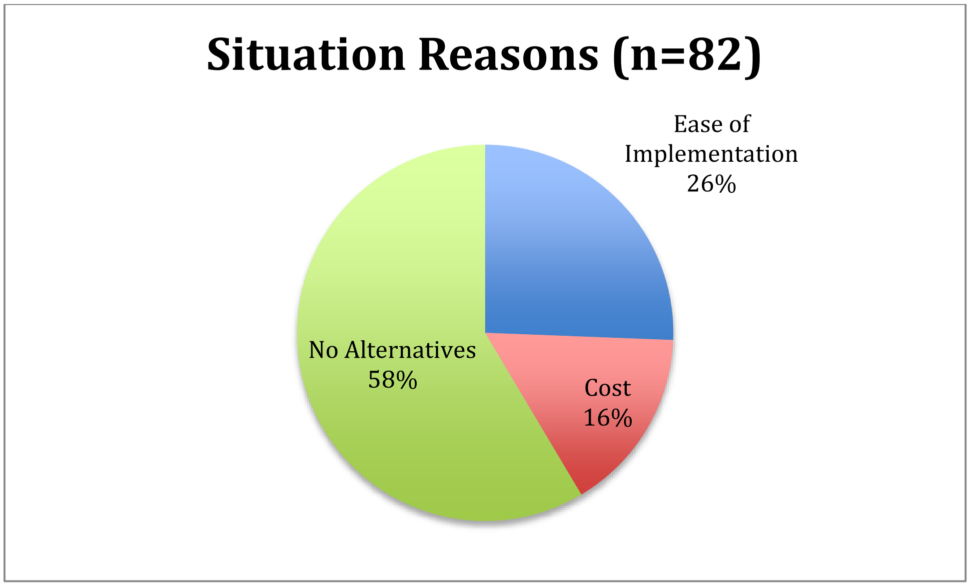


Figure 18. All Situation Reasons for actions in the SBI

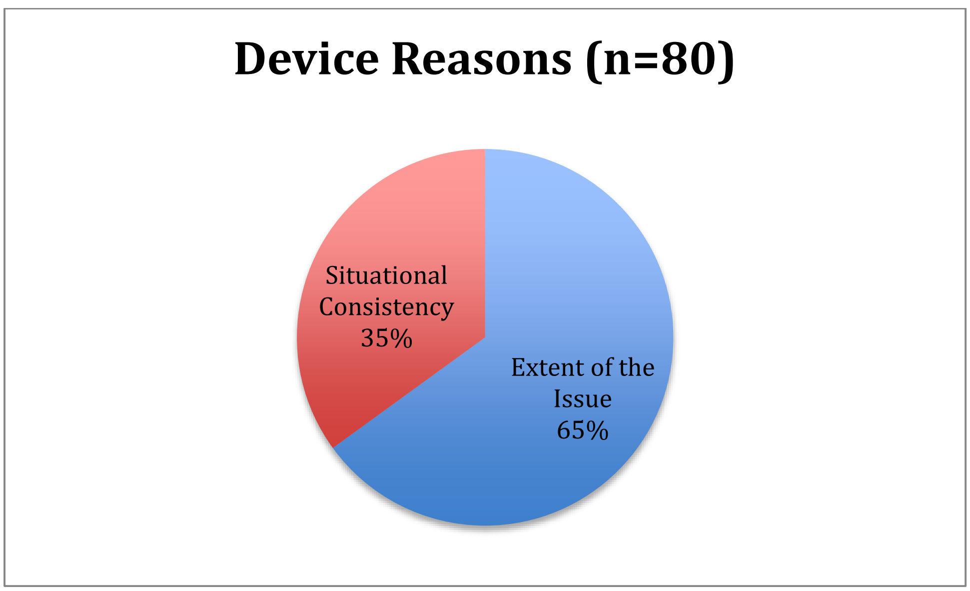


Figure 19. All Device Reasons for actions in the SBI

Although the most frequent reasons for actions varied for each scenario, the scenarios were combined to be able to examine the relationships between reasons and actions. For this analysis, any action-reason mappings that included any of the follow were eliminated: (1) an “other” action code; (2) an “other” reason code; (3) actions without a code; and (4) actions with the strategy group code of Wait For Something to Happen (note: there were no actions that only contained barrier reasons). This resulted in a total of 441 reasons and 279 actions. We conducted a χ^2 test on the total number of reasons to engage in each strategy group. The numbers of expected cues was calculated to be the same percentage of action sets. Additionally, the Wait for Something to Happen strategy group was excluded. The χ^2 test revealed that the frequency of reasons did not differ from expectations, χ^2 (4, n = 441) =4.01, p = .405. The number of reasons was not related to the strategy group, with the strategy groups with the fewest number of actions also having the fewest number of reasons. The rank order of frequency of total cues per strategies followed that of actions, except for the two most frequent strategy groups, which were very close in their number of reasons.

A chi-square test for independence between reason and strategy was not conducted because cell size was too small for Situation and Device Reasons for two of the strategy groups (and therefore also too small at the lower level coding scheme of each type of reason). However, qualitatively, the nature of the reasons and strategy group did have discernable relationships (Table 24).

The strategy groups of Change Usage Pattern (number of actions=83) and Continue to Use (number of actions=73), had fairly similar Reasons proportions to one another and to the grand total of all strategies (although these were two of the three largest categories of reasons and actions therefore they contributed greater to the grand totals). About two-thirds of the reasons for both strategy groups were Person Reasons. Person Reasons for the Change Usage Pattern trended around different types of knowledge and experiences (40 knowledge-related cues) and around the Purpose of Use/Importance of the Error (35 cues). That Knowledge was a

Table 24

Reasons for Actions in Strategy Groups in the SBI[1]

	Continue to Use (n^2=73)	Change Usage Pattern (n=83)	Gather Information or Seek Help to Get it Fix (n=76)	Change or Monitor My Behavior in the Situation (n=30)	Try to Fix it on My Own (n=17)	Total Of Each Reasons Across Strategy Groups (n=279)
			Person Reasons			
Figuring Out What Caused the Issue	3	11	14	22	12	62
Non-causal Knowledge/Experience	15	13	5	6	2	41
General Life Knowledge/Experience	7	16	37	0	2	62
Importance of Issue/Purpose of Use	37	35	6	2	6	86
Personal Preference	5	8	3	1	0	17
Awareness	8	2	2	1	0	13
Total Person Reasons	75	85	67	32	22	281
			Situation Reasons			
No Alternatives	11	16	16	2	2	47
Ease of Implementation	7	2	10	2	0	21
Cost	0	7	6	0	0	13
Total Situation Reasons	18	25	32	4	2	81

Table 24 (continued)						
	Continue to Use (n^2=73)	Change Usage Pattern (n=83)	Gather Information or Seek Help to Get it Fix (n=76)	Change or Monitor My Behavior in the Situation (n=30)	Try to Fix it on My Own (n=17)	Total Of Each Reasons Across Strategy Groups (n=279)
Device Reasons						
Situational Consistency	2	2	20	2	1	27
Extent of the Issue	20	14	16	2	0	52
Total Device Reasons	22	16	36	4	1	79
Total Reasons						
Total of Reasons in Each Strategy Group	115	126	135	40	25	441

[1] This table excludes the Wait For Something to Happen strategy group, which is also excluded in the total numbers.
[2] n is the number of action sets within each type of strategy group

frequent reason to Change Usage Pattern is not surprising given the strategy group included actions like manually entering the data into the activity tracker's log (if participants knew how to do so). Likewise, if participants found the issue to interfere with how or why the used their activity trackers, they might stop using their activity tracker all together. Continue to Use's Person Reasons was largely composed of Purpose of Use/Importance of Issue reasons (n=37). These reasons typically explained that the issue could be ignored because the participant either did not use that feature or because the participant felt the issue was not substantial enough to impact their overall ability to use their tracker for their fitness goals. Situation Reasons were similar for both strategy categories, where No Alternatives was a frequent Situation Reason, but Cost and Ease of Implementation were also mentioned to a lesser degree. Device

Reasons were also very similar with between the two strategy groups. All but 2 of Device Reasons for Change Usage Pattern and 2 for Continue to Use discussed how frequently the issue occurred or how much the automation deviated from the true value (i.e., Situational Consistency was rarely mentioned).

A second pattern of Reasons to Strategy group relationships was followed by both the Change or Monitor My Behavior in the Situation and Try to Fix It on My Own strategy groups. Both of these groups had a very large portion of their reasons as Person Reasons (80-88%). The most frequent type of Person Reason for Change or Monitor My Behavior in the Situation strategies related to Figuring Out What Caused the Issue (n=22). Participants tended to discuss how if they knew what the issue was, they might be able to do something to prevent or compensate for the issue. These causes also included instances in which the participant felt a user error had occurred. Reasons for Try to Fix It on My Own strategy groups also trended around knowing what might be causing the issue. However, most of these reasons assumed the cause was something about the device, and not user error. Situation and Device Reasons were rarely mentioned for both strategy groups. Therefore no obvious patterns were discernable regarding the nature of Situation and Device Reasons for either strategy group.

A third and final Reasons to Strategy relationship was found for Gather Information or Seek Help to Fix It. Less than half (49%) of the reasons provided for this strategy group were Person Reasons. Most of the Person Reasons related to knowledge, in particular General Life Knowledge and Experience. Participants frequently mentioned knowing that the company would be more knowledgeable then the user at fixing the issue. They also tended to discuss the importance of documenting the issue to the company and knowledge about warranties. All of these reasons explained why participants would try to contact their activity tracker's company when facing an issue. Situation and Device Reasons each made up about a quarter of the remaining Gather Information or Seek Help to Fix It Reasons. Half of the Situation Reasons (n=32) were coded No Alternative (n=16). Participants frequently explained that they would seek help (e.g., from customer service) because they had run out of ideas to try to fix the issue on their own. Ease of implementation and cost also emerged in Gather Information or Seek Help to Fix It. In particular, the actions of going online for help and searching Google, forums, and product FAQs were recommended because they were free actions and did not require a lot of effort. Device reasons (n=36) were about equally split between Situational Consistency reasons and consistency reasons of frequency and the extent of deviation from the true value. The Situational Consistency typically emerged for actions of seeking help online (e.g., through forums) to determine if other people were experience similar scenarios. However, participants explained that if the issue was re-occurring, they would instead seek help by contacting the company.

In sum, it appears that the nature of Reasons for actions are indeed related to the strategies selected for handling an automation issue. To elaborate, although Person Reasons were the largest category of Reasons, Person Reasons appeared especially salient for the Change or Monitor My Behavior in the Situation and Try to Fix It on My Own strategy groups. However, Person Reasons seemed less salient for the Gather Information or Seek Help to Fix It strategy group. Additionally, the type of Person Reasons utilized varied for different strategy groups. For instance, Person Reasons about the Purpose of Use/Importance of Issue were frequent for the Continue to Use strategy group. However, a different type of Person Reason, General Life Knowledge and Experience, was frequently reported for the Gather Information or Seek Help to Fix strategy group. One type of reason of particular interest is Person's Reasons about Figuring Out What Caused the Issue. These reasons suggest device mental models may impact troubleshooting by helping with issue interpretation and reaction. However, Reasons for strategy selection may have been skewed because the types of strategies participants reported (and there associated reasons) may have been dependent on the interview question, "Can you tell me what would you do if that did not work and the issue kept happening?" Thus, instead of Reasons, we used Issue Interpretation Details to examine the device mental models across participants.

Device Mental Models: [RS] To What Extent Do Users' Device Mental Models Relate to how They Interpret Automation Issues?

We report two scores for device mental model – one for elaborateness and one for accuracy (Table 25). The elaborateness score came from the Activity Tracker Explanation Form. There was a wide range the participants' elaborateness of device mental models, where higher scores equate to more details in the participants' device mental models. The accuracy score came from the 10 item Device Mental Model Knowledge Questionnaire, wherein higher scores represent greater accuracy of participants' device mental models. Although on average, participants had a moderately high score on this measure, participants did not typically answer all of the questions correctly. To determine if elaborateness and accuracy might be related, we correlated the first device mental model score from the Activity Tracker Explanation Form with the Device Mental Model Knowledge Questionnaire score. This correlation was not significant, $r(28) = -.17$, $p = .37$ suggesting that the two were indeed measuring different parts of device mental model (elaborateness and accuracy).

Table 25
Participants' Device Mental Models

Descriptor	**M**	**SD**	**Range**
Device Mental Model Elaborateness Score [a]	20.5	5.79	12-36
Device Mental Model Knowledge Accuracy Score [b]	7.67	1.49	4-10

Notes:
[a] Higher numbers represent a greater number of details in device mental models. Reported on the Activity Tracker Explanation Form.
[b] Scores could range from 1 to 10, with higher numbers representing greater more accurate device mental models. Reported on the Device Mental Model Questionnaire.

Because the two scores were not collinear, the scores were standardized and combined into one device mental model score. When combined and standardized, the lowest mental model z score was -1.19 and the highest was 1.5 (SD = .64). The scores created three natural groups of low device mental model scores, middle device mental model scores, and high device mental model scores (Figure 20). The number of issue details provided by participants in the SBI did not differ between groups, $F(2, 27) = 1.515$, $p=2.38$. Furthermore, chi-squared test for independent groups did not reveal any differences in the types of issue details participants provided in the SBI, and this was true for every level of issue detail (Comparing Not an Automation Error, General Automation Error, Non-causal Specific Automation Error, and Causal Specific Automation Error: χ^2 (6, N = 273) = 5.50, p =.482; Comparing Not an Automation Error and Automation Error χ^2 (2, N = 273) = .67, p =.801; Comparing General Automation Error and Specific Automation Error χ^2 (2, N = 218) = 1.21, p =.547; and Comparing Non-causal Specific Automation Error and Causal Specific Automation Error χ^2 (2, N = 154) = 3.56, p =.169). It is possible that device mental model's impact on automation issue interpretation may not have been significant because users just need a general device mental model wherein increased details may not necessarily be helpful, especially if responses would be the same for a given situation.

Although the non-significance of these test fail to provide quantitative evidence for how different device mental models impact automation issue interpretation, it still is plausible that device mental model does indeed impact error interpretation. Although cues to error interpretations were in part provided to participants in the SBI, and therefore were not segmented and coded, results from the TSI provide qualitative evidence that device mental model, such as understanding the causes of automation issues and the logic of how the device works, does impact issue interpretation. In particular, Device Mental Model emerged as a cue that helped

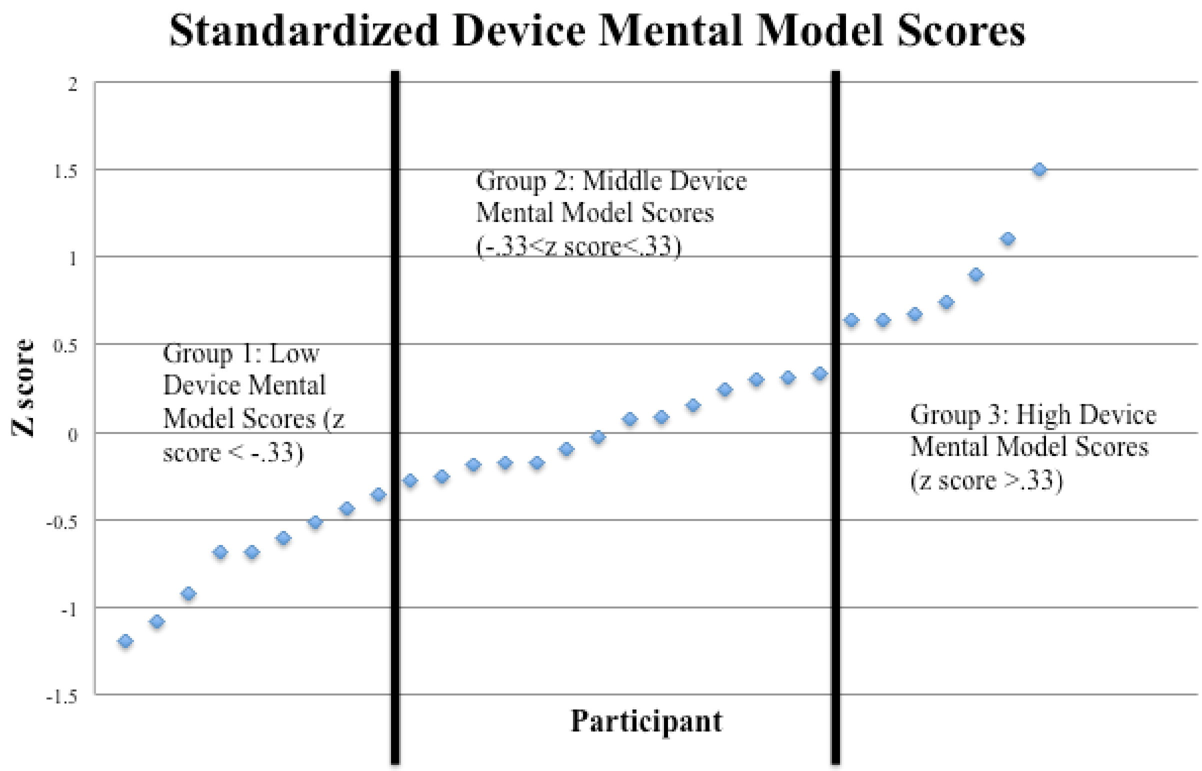


Figure 20. Standardized device mental model scores formed 3 groups.

participants classify an issue as Not an Automation Issue (e.g., when the participant understood how the device worked, including the limitations of the automations abilities). Device Mental Model, was also an important cue for Causal Specific Automation Error interpretations (e.g., when the participant understood what caused the device to function in a peculiar way.) Thus, we could not find differences across individual participants' device mental models and automation issue interpretation, device mental model still emerged in the TSI as an important cue to issue interpretation.

Summary of SBI Results

Participants generated a number of issue interpretations for each of the 5 scenarios. Although Causal Specific Automation Error was the most frequently reported issue detail (35%), it was not the most frequent interpretations in all scenarios. The strategy groups discussed by participants were qualitatively related to scenarios. For instance, the three scenarios that participants reported as being the least impactful on their ability to use their activity tracker were also the three scenarios with the greatest proportions of the Continue to Use strategy group. Likewise, Reasons also varied across scenarios, although Person Reasons were most frequently reported. However, the saliency of Person Reasons varied greatly by strategy groups, with Try to Fix It on My Own reporting 88% of its reasons as Person Reasons and Gather Information or Seek Help to Fix It only having 49% of its reasons be Person Reasons. Device mental model scores did not produce any statistically significant results as to why participants might have different automation issue interpretations. However, cues to automation issues from the TSI suggest device mental model is indeed an important factor in the troubleshooting process.

CHAPTER 6

DISCUSSION

The goal of this thesis was to gain a greater understanding of the troubleshooting process among experienced everyday automation users. In particular, this study examined the cues utilized to understand that an automation issue had occurred, the types of interpretations users make about automation issues, the reasons users select response strategies, and the types of responses users select. The major findings, theoretical contributions, and practical contributions of this study, along with its limitations and future directions are discussed below.

Summary of Key Findings

This study utilized TSI and SBI methodology to examine how experienced activity tracker users troubleshoot automation issues. Data from the TSI revealed a variety of cues to an automation issue. Some cues were rarely mentioned, such as a General Sense or Feeling, but others were frequently mentioned. The most frequently mentioned cues were context cues and measurement comparison cues. Instead of just comparing to other devices (i.e., other people in the person-perception literature), this study showed users also compare an automation's data to their own behavior, to reference points, and to other technologies demonstrate that the cues are not limited to the device itself.

This study also examined the actual interpretations users make regarding automation issues. Consistent with Consolvo et al. (2008), users did not necessarily consider limitations of the automation's design to be an automation error. Indeed, this is likely why more Not an Automation Error codes emerged in the SBI, where participants were given scenarios that could include device limitations, than in the TSI, where participants generated their own issues. However, an important finding from this study is that although it is the most frequent type of issue interpretation, users do not always try to interpret an automation issue to a causal extent. Indeed, references to device mental models, such as how the automation works, were found for every type of issue interpretation. In other words, participants reported causal cues without necessarily having a causal interpretation of the issue (although Causal Automation Error and Not an Automation did have the greatest proportion of device mental model cues). This finding suggests that sometimes automation users do not interpret an issue beyond a Non-causal Specific Automation Error, or even beyond a General Automation Error level. Lastly, participants frequently interpreted the same incident in more than one way, suggesting a certain amount of uncertainty about what was the issue.

Both the TSI and the SBI examined the reasons users select different types of strategies. The most frequent type of reason in both interviews were Person Reasons, which included various types of Knowledge and Experience, Personal Preference, and the user's reasons for using the activity tracker. Although the number of Person Reasons to carry out an action was more than double the Situation or Device Reasons in both interviews, the number of Barrier Person Reasons (n=43) was not drastically more than the number of Situation Barrier Reasons (n=37) in the TSI. Some of the Reasons and Barrier Reasons have been previously reported in the trouble-shooting or person-perception literatures, such as knowledge and experience with the activity tracker, knowing what caused the error, effort, Situational Consistency, and the Extent of the Issue (e.g., Konradt, 1995; Bereiter & Miller, 1989; Kelley, 1967; Pop 2013) and the human-automation literature (e.g., Itoh et al., 1999; Muir & Moray, 1996). However, other reasons, such as Personal Preferences, the Importance of the Issue/Purpose of Use, General Life Knowledge and Experience, Cost, and No Alternative Actions Available were not previously explored in other studies and are a unique contribution of this thesis.

Both the TSI and SBI also categorized the types of responses participants described for various automation issues. Most previous human-automation interaction studies have only allowed minimal response choices for participants (e.g., rely on the automation more, rely on the automation less; Riley et al., 1996). Additionally, the troubleshooting of computer studies had also limited strategies to focus on the actual end result of fixing the issue (Konradt, 1995; Bereiter & Miller, 1989). However, in everyday life, users have more choices. The present study revealed a rarely used strategy of Wait for Something to Happen. However. However, greater contribution of this study is the three prevalent categories of strategies of Gather Information or Seek Help to Get It Fixed, and Change or Monitor My Behavior in the Situation, and Try to Fix It on My Own (although fixing the issue is in line with trouble-shooting studies; Konradt, 1995; Bereiter & Miller, 1989). Both of these groups of strategies view the user as an empowered agent that can do more than simply use or not use the technology. Change Usage Pattern was the most frequently reported strategy in both the TSI and SBI, but this could have been the result of asking participants what would happen if their previously mentioned strategies failed (to which participants would frequently reply that they would stop using the activity tracker, or replace it). However, the unique strategy groups of Gather Information or Seek Help to Get It Fixed, and Change or Monitor My Behavior in the Situation, and Try to Fix It on My Own were sizable in both interviews.

The current study also attempted to address a supplemental research question of the role of device mental model in automation issue interpretation. Although differences in issue interpretation were not detected from the measures of device mental models, it is still clear that device mental models play some sort of role in issue interpretation. This was evident from the qualitative data on cues to issues, where device mental model emerged as a frequent cue. Future research should further examine what exactly is the role of device mental model in troubleshooting.

The greatest contribution of this study, however, is the relationship between the different types of data. Specifically, we mapped cues to issues with issues interpretation and mapped reasons for strategies to strategies. The originally proposed theoretical model was revised to incorporate these findings. The revised troubleshooting models are discussed next in the theoretical contributions.

Theoretical Contributions

Prior to this study, there were no theoretical frameworks for the troubleshooting process of everyday automation issues. In the introduction, this thesis synthesized the previous literature and proposed a guiding theoretical framework. Several major revisions to the framework occurred based on the findings of this study. First, because participants sometimes interpreted the same incident multiple times, we could not link directly from issue interpretation to reasons for actions and instead split up the troubleshooting model in Automation Issue Interpretation and Automation Issue Reaction. Secondly, the revised model includes automation issues, instead of just automation errors. Findings suggested classifying the issue as Not an Automation Error could be an involved process with causal cues, and responses to Not an Automation Error are likely not limited as originally thought to the Continue to Use strategy group. Additionally, device mental models were incorporated as a cue that an issue had occurred and the types of strategy groups were replaced by the groups that actually emerged in the study (which were not clear-cut into general, specific, or causal strategies and were better categorized by natural groupings). Finally, details about the types of cues to issues and reasons to strategies were also incorporated.

Issue Interpretation

Figure 21 shows a revised model of the issue interpretation part of troubleshooting and is based on the TSI results. On the left, the 7 cues to an automation issue are presented in order of frequency (note: we eliminated the codes that made up less than 5% of the cues). The second column starts the different categories of issue interpretation. This study did not address cues that are not attended to, but acknowledges situations exist wherein cues could go unattended. For each type of issue interpretation, the three most frequently reported cues are listed, unless multiple types of cues tied in frequency (in which case all in the tied 3rd most frequent cues are reported). Notably, Information Provided About the Device and Component Information did not obtain a great enough frequency to map to any issue interpretations. Context, Measurement Comparison, and Device Mental Model mapped onto most of the interpretations. Check Device cues were mapped onto the decision of if the issue was an automation error or not and onto determining Non-causal Specific Automation Error interpretations. Consistency mapped onto 3 different interpretations: An Automation Error, General Automation Error, and Non-causal Specific Automation Error. In short, Figure 21 provides a summary of the major cue-to-issue-interpretation relationships.

Figure 21 builds greatly on the human-automation and person perception literatures. First, it supports the cues hypothesized in the literature review. In particular, the Context cues of knowing where, when, or on which tasks an automation might err, are frequently reported. This suggests the findings of Madhavan et al. (2006) and Masalonis (2003) generalize to automation in everyday life. Further, the prevalence of Context, Measure Comparison, and Consistency in automation issue interpretation suggests that constructs from human-human interactions can translate to human-automation interactions (Pop, 2013). Indeed, although the second most frequent cue was Measurement Comparison, that cue was only found in the person-perception literature (Kelley, 1967; Pop, 2013). A theoretical contribution of this thesis is that Measurement

Comparison is a prominent cue in automation issue interpretation and should be furthered study, perhaps through experimental manipulation, as a cue used in issue interpretation.

Another contribution of this study was the support for the role device mental model can have in automation issue interpretation. In particular, the emergence of Device Mental Model cues as the third most frequent type of cue supports the generalizability of Lees and Lee (2007)'s results to other automation (from driving alerts to activity trackers). Further, although not especially frequent, the emergence of Checking Device as a frequent cue provides another theoretical contribution. Often, human-automation interaction studies treat verifying the automation as a dependent measure (typically measuring the construct of "use" or "trust"; e.g., Masalonis, 2003). However, the results of this thesis suggest the amount that a user checks their automation might be helpful in understanding if an issue is actually an automation error and in classifying where the error is or what is impacted by the error. Further investigation into checking automation might explain why it emerged as a cue – for instance, perhaps checking the automation is helpful in conjunction with Context cues, to understand which situations do and do not seem susceptible to error.

Consistency, Information Provided About the Device, and Component information also emerged in this study and were derived from the human-automation literature. For instance, both the timing (e.g., Itoh et al., 1999) and amount of variability of the error (Muir & Moray, 1996) have been studied in relation to Consistency (frequency). However, it appears that Consistency often did not help with the more casual explanations. Instead, Consistency appears to help more with recognizing that an issue has occurred, or even where it is occurring, but not why. Interestingly, Information Provided about the Device did not frequently emerged for Causal-Specific Automation Errors either. This suggests that studies that simply tell users why an automation might err (e.g., Bisantz & Seong, 2001; Dzindolet et al., 2003) might be limited in generalizability. In short, the causal information provided to users may not actually be the same cues they use when interpreting automation issues on their own. However, this lack of generalizability may depend on the amount of training with the automation. Future work should test if training material is more frequently cited for more complex automation, and less so for everyday automation.

Within cues to automation issues, some cues were not particularly frequent. Component information rarely arose, although it was hypothesized to play a role (Rovira et al., 2014). However, perhaps this also depends on the type of automation. It is possible that more complex automation, or automation that assists more with decision making as opposed to sensing or detecting, might influence the user to rely on Component Information more so. Additionally, other cues were not even included in Figure 21 because they were extremely rare. For example, General Sense or Feeling was less than 2% of all cues. This low frequency, along with the other cues types that were frequent, suggests that users attempt to ground their interpretation in facts, as opposed to just hunches.

This study also contributes to the actual interpretations of automation issues. First, our study supports the findings of Consolvo et al. (2009) in that users may not classify all automation issues as errors, especially if they have cues from their device mental model. Further, the designs of previous studies have made it difficult to know what the user was actually thinking. For instance, it is unclear in the study of Itoh et al. (1999) why difference usage patterns emerged between participants that received errors in a series followed by a long period of no errors and participations that received errors spread out equally over time. The findings of this study suggest that the participants may not have been coming up with causes for the issue. Rather, they may have simply accepted that something was wrong. In short, this study suggests that just because users *might* interpret an issue as having a particular cause does not mean that they *will* interpret an issue as having a particular cause. Additionally, although we did not code for accuracy, participants sometimes did provide explanation that were factually incorrect. Thus, the explanations users come up with, if they come up explanations, may be erroneous (e.g, Chapman & Champan, 1969).

Because no previous study has explored what users actually thought the automation issue was, previous studies have also not been able to map cues to interpretations. The mapping that we provide is a good first step in understanding how to shape user's interpretation. For example, simply having an automation issue occur frequently is not necessarily helpful for the user to understand what caused the issue. Rather, Context and Device Mental Model help shape causal interpretations. Further, the SBI provided additional unique insight that is not captured in Figure 21. In particular, the SBI demonstrated that the concreteness of the situation impacts how the user interprets the issue. In particular, users are more likely to come up with explanations for an issue or understand what parts of the automation are impacted by the issue when the issue is more concrete. In contrast, in Scenario E wherein the device was unresponsive and the situation was vague, participants did not as frequently attempt to suggest a cause. Rather, participants frequently only provided a General issue interpretation. Although we were unable to link interpretation directly to reaction, perhaps the impact of concreteness is because users may not invest cognitive resources into guessing why the issue occurred if their reaction would be the same regardless of the cause (e.g., stop using the unresponsive tracker). Future work should examine if and how interpretations directly link to reactions.

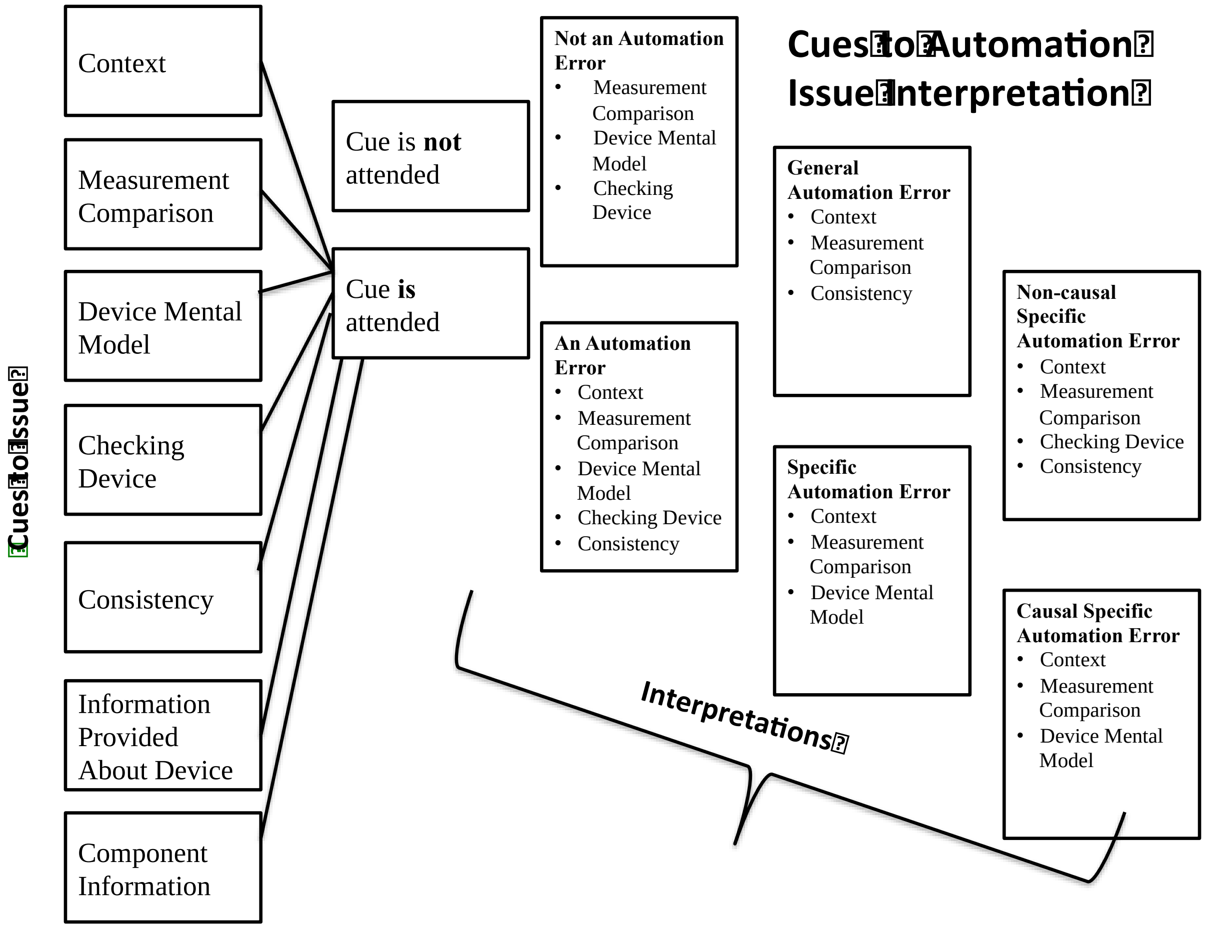


Figure 21. Conceptual model of automation issue interpretation during troubleshooting based on data from the TSI. Cues that were 5% or more of the total cues are included as cues that may be attended too. Cues are order in descending order of frequency. The Issue Interpretation coding scheme follows in the three right-most columns. For each interpretation, the three most frequently mentioned cues are listed, unless cues tied in frequency.

Issue Reaction

Figure 22 shows three revised models of the issue reaction part of troubleshooting and is based on the TSI and SBI results. For simplicity and clarity, Reasons are divided into Person, Situation, and Device Reasons across Figures 22a, 22b, and 22c. The five most frequently mentioned groups of strategies are in the in the middle with examples of the strategies within each of the models. The Wait for Something to Happen strategy was not frequently used and is not described here. Strategies are ordered form left to right by their total (SBI + TSI) frequency.

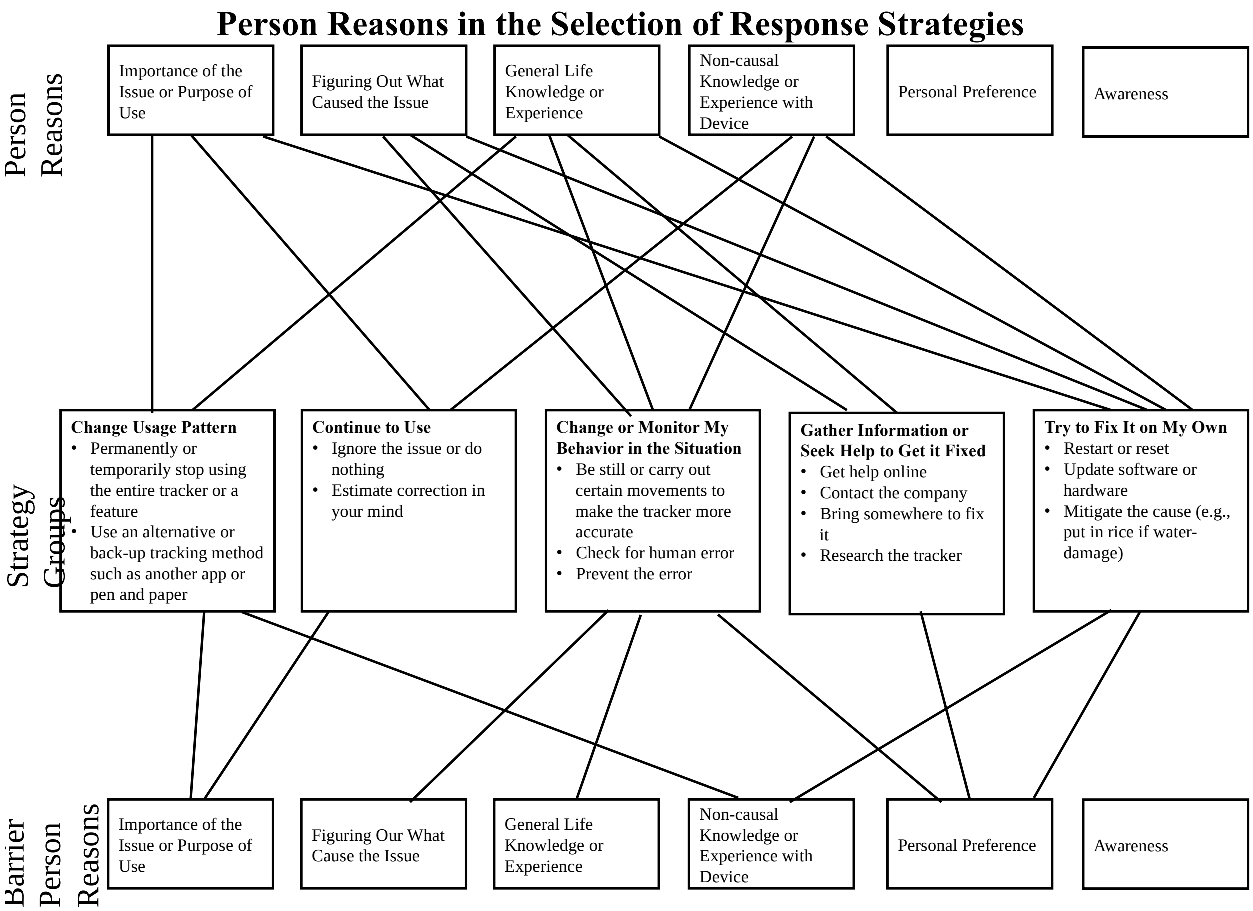


Figure 22a. Conceptual model of Person Reasons and Person Barrier Reasons, and strategy groups for responding to automation issues based on TSI and SBI data.
Figure 22. Conceptual model of strategy selection when troubleshooting automation issues based on combined TSI and SBI data. Person, Situation and Device Reasons have been separated out. Strategy groups are listed from left to right in order of total frequency. Examples of actions for each strategy group are included. Reasons are listed from left to right in order of total frequency. Barrier Reasons are listed in the same order as Reasons and not necessarily in order of frequency. Barrier Reasons only include data from the TSI. Mappings between Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Reasons for the strategy group when TSI and SBI data were combined. Mappings between Barrier Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Barrier Reasons for the strategy group and if they included more than one Barrier Reason in TSI data.

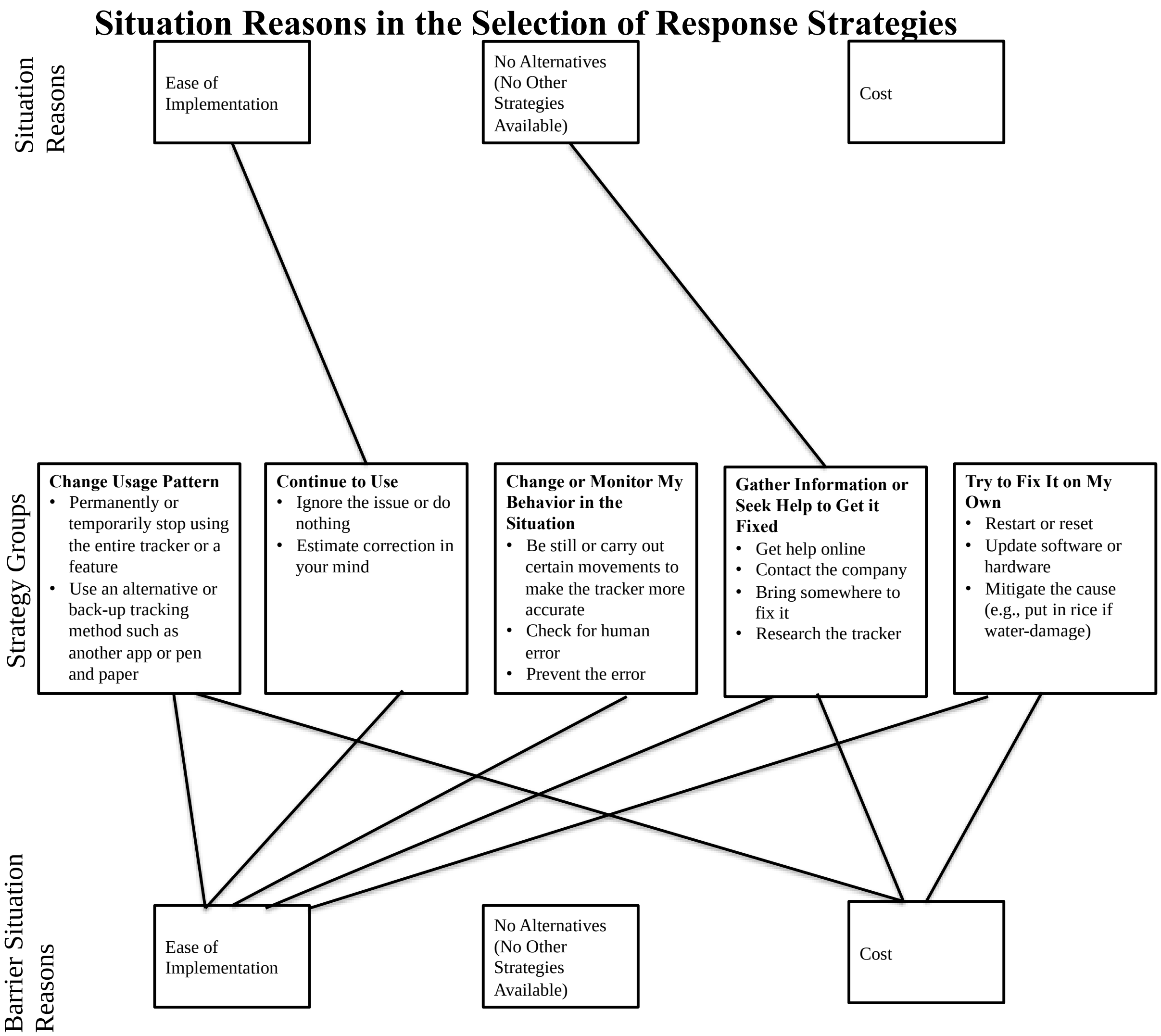


Figure 22b. Conceptual model of Situation Reasons and Situation Barrier Reasons, and strategy groups for responding to automation issues based on TSI and SBI data.
Figure 22 (continued). Conceptual model of strategy selection when troubleshooting automation issues based on combined TSI and SBI data. Person, Situation and Device Reasons have been separated out. Strategy groups are listed from left to right in order of total frequency. Examples of actions for each strategy group are included. Reasons are listed from left to right in order of total frequency. Barrier Reasons are listed in the same order as Reasons and not necessarily in order of frequency. Barrier Reasons only include data from the TSI. Mappings between Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Reasons for the strategy group when TSI and SBI data were combined. Mappings between Barrier Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Barrier Reasons for the strategy group and if they included more than one Barrier Reason in TSI data.

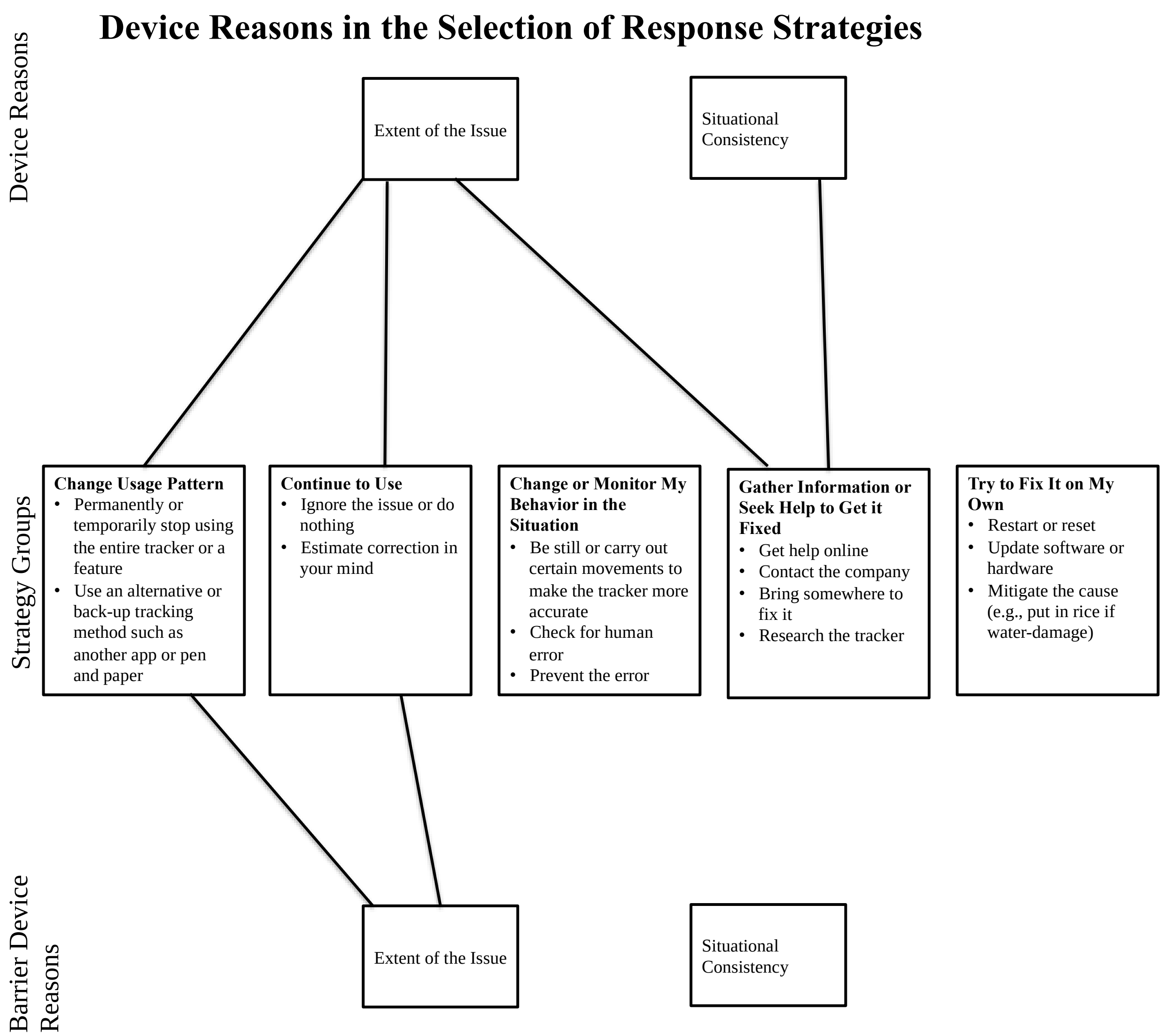


Figure 22c. Conceptual model of Device Reasons and Device Barrier Reasons, and strategy groups for responding to automation issues based on TSI and SBI data.
Figure 22 (continued). Conceptual model of strategy selection when troubleshooting automation issues based on combined TSI and SBI data. Person, Situation and Device Reasons have been separated out. Strategy groups are listed from left to right in order of total frequency. Examples of actions for each strategy group are included. Reasons are listed from left to right in order of total frequency. Barrier Reasons are listed in the same order as Reasons and not necessarily in order of frequency. Barrier Reasons only include data from the TSI. Mappings between Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Reasons for the strategy group when TSI and SBI data were combined. Mappings between Barrier Reasons and Strategy Group are only included if the Reason was 10% or more of the total number of Barrier Reasons for the strategy group and if they included more than one Barrier Reason in TSI data.

For each strategy group, if the type of reason composed 10% or more of the number of Reasons or Barrier Reasons for that strategy group when TSI and SBI data were combined, and 10% was greater than 1 Reason or Barrier Reason, it is mapped in the diagram. Reasons are ordered by their total frequency and Barrier Reasons are ordered the same as Reasons.

This thesis contributes to the literature on human-automation interaction by classifying reactions that have previously been ignored by the human-automation interaction literature. Almost all human-automation interaction studies only allow the users to continue to use the automation or to stop relying on or trusting the automation (e.g., Dzindolet et al., 2003; Madhavan et al. 2006;,Riley, 1996). This study suggests that when constraints are removed from participants, they may respond to automation issues with a greater variety of action. The strategy groups unique to this study include Change or Monitor My Behavior in the Situation, Gather Information or Seek Help to Get it Fixed, and Try to Fix t on My Own. Future research should lighten the constraints on participants to determine if these unique strategies generalize to other types of automation.

Likely because of the limited reactions studied in the human-automation literature, previous work in the human-automation literature and the trouble-shooting has not to consider reasons why users may not take a particular action (e.g., Bereiter and Miller, 1989; Konradt, 1995). A theoretical contribution of this thesis is the beginning of studying Barrier Reasons in human-automation interaction. Although the number of Barrier Reasons was small, Person and Situation Barrier Reasons were frequently reported for all strategy groups compared to Device Barrier Reasons. Of the Person Barrier Reasons, Personal Preference, which is only a Barrier Reason in the model (Figure 22a), was reported for the three unique strategies of Change or Monitor My Behavior in the Situation, Gather Information or Seek Help to Get it Fixed, and Try to Fix it on My Own. In contrast, Importance of the Issue or Purpose of Use Person Barrier Reasons were only frequently reported for the previously documented strategy groups Continue to Use and Change Usage Pattern. For Situation Barrier Reasons, Ease of Implementation emerged for every strategy group. This is consistent with from Konradt (1995) and Bereiter and Miller (1989) results that effort is a factor in selecting a reaction (i.e., high effort makes selection of the strategy less likely). Device Barrier Reasons were about the Extent of the Issue and were only frequently reported for the previously documented strategy groups of Continue to Use and Change Usage Pattern. Although this study provides theoretical contributions with Barrier Reasons, the total number of barrier reasons should be considered a limitation of this study, and also as a point for future research.

This thesis builds on the troubleshooting studies of from Konradt (1995) and Bereiter and Miller (1989) and provides evidence for the generalizability of some of their findings. In addition to effort, reasons relating to knowledge, such as the knowledge and experience that technicians had used in troubleshooting large computer systems in both studies, were also found for users of activity trackers. Indeed, the results of this thesis suggest that knowledge (General Life Knowledge or Experience, Non-causal Knowledge or Experience with the Device, and Figuring Out What Caused the Issue) is frequently a reason why users try a strategy. However, this thesis goes beyond the work of from Konradt (1995) and Bereiter and Miller (1989) in that it considers the nuances between knowledge types and strategy groups. For instance, Figuring Out What Caused the Issue was frequent for more pro-active strategies such as Change or Monitor My Behavior in the Situation, Gather Information or Seek Help to Get it Fixed, and Try to Fix it on My Own compared to Change Usage Pattern or Continue to Use. For Device Reasons, although Barrier Reasons demonstrated the importance of Ease of Implementation (Konradt, 1995;

Bereiter and Miller, 1989), Ease of Implementation only emerged as a Reason for the Continue to Use strategy. This result suggests that the findings of Konradt (1995) and Bereiter and Miller (1989) need to be contextualized within the actions the user takes. Both previous studies also found that historical information, such as the frequency of the issue (i.e., Extent of the Issue) was a Reason technicians gave in troubleshooting. This thesis also found Extent of the Issue as a Reason for multiple strategies and suggests that that finding can be generalized to other automations and to users with minimal training.

Methodological Contributions

Although the TSI and SBI interviews typically complemented one another, both provided different value to this thesis. The TSI was valuable in that it provided details from automation issues actually experienced by users and issues that were salient enough from them to be remembered. Indeed, likely because participants were asked to generate issues in the TSI, a smaller proportion of the TSI interpretation details were Not an Automation Error (13% in the TSI; 20% in the SBI). The TSI was also helpful in revealing cues to issues and Barrier Reasons to Strategies. However, the TSI was more focused on why to try an action as opposed to why not. Future research could build on this method by focusing more on reasons why people might not respond to automation issues (e.g., explicitly ask “are there any reasons why you might not try that strategy”.) The SBI was helpful in examining across typical issues users might incur and for controlling the data more (e.g., participants in the TSI generated differing numbers of incidents whereas all participants received 5 incidents in the SBI). The SBI also demonstrated that the situation, in particular the impact of the situation, relates to the strategies attempted by a user. Depending on the circumstances, either interview method could be employed. For example, if a designer wanted to know what cues to issues users rely on, the TSI method would be more appropriate. However, if a researcher wanted to know details about a particular situation, or if a researcher wanted to directly link interpretation to reaction, the SBI, with issue interpretation details, would be more appropriate.

Additionally, the SBI provided a testing ground for the device mental model supplemental research question, although no significant results were found. Given the device mental model questionnaires were developed for this study, insignificant results might have occurred due to necessary refinements in the device mental model measurements. First, it might be the case that both accuracy and elaborateness are not equally relevant in issue interpretation. Second, it might be the case that the elaborateness construct was over-simplified. It could be that providing fewer details about the device is actually indicative of more elaborate device mental models wherein “chunking” of detailed features into larger categories has occurred and thus the larger categories were so engrained that they were the only information participants described. Finally, it could also be that the measurements simply were not valid. For instance, perhaps certain items on the accuracy measurement varied more than anticipated for each tracker and additional details for each item may have been necessary for accuracy scores. Alternatively, more detailed directions in drawing out how the activity tracker works might have promoted more detailed responses from participants for elaborateness scores. Future work could attempt to refine device mental model measurements or determine a more standardized way to measure device mental models as a whole (e.g., relatedness ratings in conceptual maps).

Practical Contributions

This study focused more on the perceptions of users than the accuracy of the users interpretations. However, the multiple interpretations provided by participants suggested that they might have been uncertain about what was the correct interpretation, and sometimes participants gave answers the researcher knew to be incorrect. However, determining what users think on their own is an important first step to ensuring accurate interpretations when necessary. In other words, by knowing what cues help to lead to certain interpretations, designers can make the correct cues salient when accuracy matters. For instance, one could imagine that knowing that an issue was occurring because of a software update might be important to be accurately interpreted (e.g., as opposed to interpreting it as a permanent issue and ceasing use of the activity tracker). In that case, designers may want to facilitate a device mental model that will be a cue to the correct Causal Automation Error interpretation. For instance, designers might want to provide notifications when new updates become available, as many technologies already do. Designers may also want to encourage their users to develop a vague understanding of how the software works—for instance, the hardware on the activity tracker records movement, but the software analyzes that data to determine what counted as a step. In this way, if a user does take the software update and the issue does not go away, the user might pinpoint the issue to the device hardware. In keeping with the software example, software updates were a Try to Fix It on My Own strategy. People often selected Try to Fix It on My Own strategies because of knowledge as to what caused the issue. This furthers supports the notion of making software updates, and the issues they fix, salient to users and suggests that technologies that do so follow a best-practice.

Sometimes users may not know or care about the cause of the issue. In these cases, it may not be necessary or may not be feasible to support the user's knowledge about what the issue is, and rather it may just be feasible to support what the user should do. For instance, data from this study demonstrate that users often seek help for issues that happen frequently or across different situations. Automation designers could help draw an individual's attention to such issues. For example, if a Fitbit HR user manually adjusted the activity recordings (i.e., 500 calories burned) and entered in the activity (e.g., spin-class) multiple times, the activity tracker could flag the instances and provide a pop-up to report an issue or lead to an FAQ. In that case, the Fitbit HR should have picked up on the spin class because in measures heart rate. However, a different technology that does not measure heart rate, such as the Jawbone Up 24, might want to encourage users to select the Continue to Use strategy because the device is not designed to be able to measure motion without arm movements. In that case, the designers might want to focus more on Person Reasons, such as promoting the user's knowledge of how the technology works and emphasizing that the purpose of using the tracker might be for overall health, rather than a particular class.

Additionally, in cases where the user may now know what caused the issue, alternative help methods might be necessary. If it is a vital situation for the participant to understand the issue, the results of this study suggest another possible way to organize help manuals is through their cues to automation issues instead of interpretation details. For instance, instead of a "Sensitivity Setting" indexed term, a term representing the discrepancies in distances between a known distance and the activity tracker's reading might be helpful (e.g., an FAQ term of "Why does my activity tracker give a different distance than my GPS?" Users could then use these cues to understand the issue better. Future research should try to link interpretations to responses. For instance, by gaining an answer to the FAQ that the sensitivity setting is wrong and can be changed, users are empowered to change the setting and fix the issue. With other response strategies, it may just be matter of letting the user know that those responses (e.g., changing input) exist. Additionally, Barrier Reasons results suggest that those responses should be easy to implement. For instance, help could be made easier by having contact information readily accessible to the user during the issue. In short, the results of this study allow specific recommendations for training, help manuals, and designers depending on the optimal reaction from the user.

Limitations and Future Directions

Like all studies, the limitations of this thesis should be considered. Post-event interview methods are subject to participant biases and memory. Additionally, although we tried to avoid leading questions, asking participants about the incidents may have resulted in further interpretation than would have occurred outside of the study. Although inter-coder agreement was high, there is a possibility that reaching reliability early on may have resulted in missing salient themes. However, given few segments were classified as “others,” it appears that coders likely did not miss an additional category. It is also worth nothing that the interview methodology is also a strength of the study because it allowed for rich details to emerge about the troubleshooting process.

Because participants could provide more than one issue interpretation per incident or scenario, we were unable to close the gap between issue interpretation and issue reaction. However, future research could do so by giving participants an interpretation and asking them how they would respond. Another limitation of this study is that the data only allow us to make conclusions about which cues, interpretations, reasons, and strategies are most frequent, not most important. Indeed, although we did not consider the order of strategies, the least desirable ones (e.g., stop using the activity tracker) tended to occur after other strategies were exhausted, suggesting the frequently mentioned strategy group of Change Usage Pattern may not be the most important strategy group. Future studies could ask participants how they would rank different cues, reasons, and strategies to determine which are cues, reasons, and strategies are the most important.

Although we had a variety of different activity trackers in this study, there is a need to generalize these findings to other automations. The ability to do so seems promising because many of the cues in the human-automation literature and the trouble-shooting literature were also found in this study. In particular, future studies should attempt to generalize to other everyday technologies that might have a greater range of potential strategies than other automations typically study (i.e., such as the expensive, not everyday automations, of luggage-screeners and chemical-processing plants).

As research into human-automation interaction continues, it is imperative that the field considers how experienced users handles automation issues, especially automation errors, in their everyday lives. Doing so will allow for a more appropriate breadth in capturing the cues people use to determine an issue and the strategies they choose to implement. Human factors interventions that focus on highlighting critical cues as reasons could aid in appropriate strategy selection. As automation, and automation imperfections, continue to become even more widespread and integrated in daily life, it will be even more critical to establish optimal human-automation interactions.

PPT ID _____

APPENDIX A

WELLNESS MANAGEMENT TECHNOLOGY BACKGROUND QUESTIONNAIRE

A wellness management technology is a technology you use to keep track of your fitness, health, or over-all wellbeing. Your activity tracker is an example of a wellness management technology.

This questionnaire is designed to assess your background with wellness management technologies.

1.What is the make (e.g., Fitbit) and model (e.g., Flex) of the activity tracker you use?

Make ___________________ Model________________________

2. What is the type of phone (tablet, computer or other device) you typically view your activity tracker data on (e.g., iPhone5, Samsung Galaxy S6)? If you use multiple devices please list all (e.g., iPhone 5 and macbook).

3. If you have your phone with you, please look up the operating system software it currently uses and write it here:

4. How many days per week do you typically wear your activity tracker?

☐ Less than 1 day a week

☐1 day a week

☐2 days a week

☐3 days a week

☐4 days a week

☐5 days a week

☐6 days a week

☐7 days a week

5. How frequently do you check any part of your activity tracker? This includes checking the app, the website, or the wearable device itself.

☐Less than once a day

☐Once every 16 hours

☐Once every 12 hours

☐Once every 8 hours

☐Once every 4 hours

☐Once or more every hour

6. How long have you been using your activity tracker?

☐Less than 1 month

☐1 month

☐2 months

☐3 months

☐4 months

☐5 months

☐6 or more months

7. Did you use a different wellness management technology before using the your activity tracker? Some examples of wellness management technologies include: Jawbone Up 24, Fitbit One, Nike + Fuel Band SE, myfitnesspal.com, and pedometers.

☐ YES ☐ NO

   a. If yes, which technology/technologies? Please list all wellness management technologies you previously used.

   b. If yes, for how many **months** did you use each of these technologies?

PPT ID _____

8. Were you keeping track of your fitness prior to using any wellness management technology?

☐ YES ☐ NO

a. If yes, how were you keeping track of your fitness prior to using any wellness management technology?

☐ In my head

☐Paper

☐A spreadsheet

☐A medical device

☐Other

PLEASE LIST: ____________________

9. Why did you choose to use your activity tracker?

10. How much do you use your activity tracker for fun? (please check one)

☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7
Not at all for fun						Completely for fun

11. How much do you use your activity tracker to manage your health and wellness? (please check one)

☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7
Not at all for health and wellness						Completely for health and wellness

12. Overall, do you think your activity tracker has a tendency to:

☐ Underestimate

☐Overestimate

☐Both underestimate and overestimate

☐Neither underestimate or overestimate

13. What percentage of the time do you find the automated data collection on your activity tracker to be correct?

☐ 0-10 %

☐11-20%

☐21-30%

☐31-40%

☐41-50%

☐51-60%

☐61-70%

☐71-80%

☐81- 90%

☐91-100%

PPT ID _____

APPENDIX B

THREAT STRATEGY INTERVIEW SCRIPT

<Critical Incident/TSI>

In the next part of the interview I want to learn more about issues that might occur with the {activity tracker} and what you would do about them.

Stage 1: Introduction of the critical task and elicitation of past experiences with critical task

1. Many people use an activity tracker with the goal of keeping track of their activity, fitness, or health. Can you tell me how someone would go about doing that?
    a. Prompt: how would they go about using it?

Stage 2: Elicitation of threats and cues to threats

Okay, so now I would like you to think of issues that could make it difficult for you to use the {activity tracker} to keep track of your activity, fitness, or health accurately. I would like for you to follow along with me here as we define an issue with the {activity tracker}.

** Think about times when your {activity tracker} was supposed to do something for you, but you have think it did something wrong.

- For example, maybe the {activity tracker} provided information that differed from what you thought it should say under those circumstances.
- Or, your {activity tracker} may have made a mistake with sensing, detecting, information processing or any thing the {activity tracker} does on its own.

This might include situations where it made a very particular error on a feature or you thought something was going on, even if you did not know exactly what is was.
**

9. Can you give me some examples of some situations like this that have or could happen to you?

So, now I would like to talk about each one of those in more detail. {repeat the following questions for each of them; but when you go through them a second time you do not need to repeat the strategy definition}

10. Can you tell more about the nature of the issue {or describe which issue you are talking about}?
    a. Prompt: What do you think may have been going on?
11. How do you become aware of the issue?
    a. Prompt: Why would you think that issue occurred?
    b. Prompt: What signs are there that issue occurred?
12. **If *error is specific*:** Why did you think that [*repeat the cue(s) they mentioned*] was related to [specific problem]?

    OR

    If *error is vague*: Why did you think that [*repeat the cue(s) they mentioned*] was a problem with your {activity tracker}?
13. When did this issue occur?
14. Can you tell me what happened just before this issue occurred?

Stage 3: Elicitation of strategies and cues to strategies

Now I will ask you about the strategies you might have used or could use to keep that issue from interfering with your use of the {activity tracker} to track your activity, fitness, or health.

I would like for you to follow along with me here as I explain what we mean by a strategy.

** A strategy is a plan or method to achieve a goal. A strategy is not usually one action, but you may think of it as an action.**

15. Can you tell me a strategy you might use to keep that issue {or describe the issue} from interfering with tracking your activity, fitness, or heath?
16. Can you tell me why you would try that particular strategy {or, if multiple strategies, describe the ones you are talking about}?
17. ***Only if they have an intervening strategy****:* How can you tell if {repeat actions or strategy A} is working or is not working?
18. What would you do if {repeat actions or strategy A/that} did not fix the problem?
19. Why would you choose that particular strategy {strategy B}?

Stopping rule: discontinue use permanently, not to track, to track wrong, to send back to manufacture, to manually override the automation, or to track wrong and correct in your mind.

20. Did this situation actually occur?
    a. *If yes,* what did you actually do?
    b. *If yes*, what was the outcome of what you actually did?
    c. *If yes*, did you do anything else in response? Maybe later on?
21. Now that we have talked about all the strategies you actually used or could have used, I would like to know more about what would make you choose {the first actions or strategy A } over {the second actions or strategy B} or the other way around. Could you think about a specific piece of information or a cue that would encourage you to choose {strategy A} over {strategy B}? * Include for any strategies the actually did but didn't explain.
22. Were there any previous experiences you have had that was helpful to you in thinking

about how you could deal with this situation?

PPT ID _____

APPENDIX C

IMPACT AND CONFIDENCE SCALES

What would be the **impact** of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale?

1	2	3	4
no impact	minor impact	moderate impact	serious impact

PPT ID _____

How **confident** are you in your judgment about this situation?

1	2	3	4	5	6
extremely confident	moderately confident	slightly confident	slightly confident	moderately confident	extremely confident
it is NOT an issue	it is NOT an issue	it is NOT an issue	it IS an issue	it IS an issue	it IS an issue

PPT ID _____

APPENDIX D

SCENARIO BASED INTERVIEW SCRIPT

So keep thinking about things the {activity tracker} does for you that you do not have to do.

**Please take a moment to look over these definitions again **

In this this next part of the interview, I am going to describe some scenarios to you about things that may happen with your {activity tracker}. For each one I will ask you a series of questions about what you think and what you might do in that situation. As a reminder, some of my questions will seem repetitive, so it is okay if your answers overlap. And again, I am trying to learn from you, this is not a test, there are no right or wrong answers. We are going to be using a couple of scales in these scenarios, so let us take a moment to go over them now.

This is the first scale <give to participant> , and it will be used to answer how much of an impact the scenario has on your ability to track your activity, fitness, or health using your activity tracker. Notice it has 4 points. 1 means the situation describes in the scenario has no impact. 2 means it has a minor impact. 3 means it has a moderate impact. 4 means it has a serious impact.

This is the second scale <give scale to participant> and it will be used to answer your confidence in how you judge the scenario as an issue or not an issue. Notice it has 6 points. 1 means extremely confident it is not an issue. 2 means moderately confident it is not an issue. 3 means slightly confident it is not an issue. 4 means slightly confident it is an issue. 5 means moderately confident it is an issue. 6 means extremely confident it is an issue.

<Scenarios>

A. Imagine you have just completed a rigorous indoor spin class. However, your {activity tracker} has not substantially increased the number of calories you have burned.

1. What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale? Note that this is a 4 point scale.
2. Can you please describe what you think is going on in this scenario?
3. Do you think this indicates an issue with the {activity tracker}?
    a. How confident are you in your judgment of this situation (that is whether or not this is an issue). Where 1=extremely confident it is not an issue and 6=extremely confident it is an issue.
        i. *If yes (4 or greater)*, what do you think the issue is?
            a. *If error that is specific or has a cause,* what would suggest that was the root of this issue?
                i. Prompt: As is in what signs are there that {insert what participant says} is related to this problem?
        ii. *If No (3 or below),* Why do you not think this indicates an issue with the {activity tracker}?
4. How would you react to this situation to be able to track your activity, fitness, or health?

<continue if the response does not include a why>

    i. Prompt: What would you do and why?
    ii. Prompt: what would make want to try that strategy for coping with this situation?

5. *If applicable*, Can you tell me what would you do if that did not work and the issue kept happening?

<if the response does not include a why>

i. Prompt: what would make want to try that/those/these strategy for coping with this situation?

B. Imagine you have just walked all of Georgia Tech's Pi Mile trail, which you know to be exactly 3.14 miles. You notice that your {activity tracker} says you have walked 3.5 miles.

1. What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale?
2. Can you please describe what you think is going on in this scenario?
3. Do you think this indicates an issue with the {activity tracker}?
   a. How confident are you in your judgment of this situation (that is whether or not this is an issue). Where 1=extremely confident it is not an issue and 6=extremely confident it is an issue.
      i. *If yes (4 or greater)*, what do you think the issue is?
         a. *If error that is specific or has a cause,* what would suggest that was the root of this issue?
            i. Prompt: As is in what signs are there that {insert what participant says} is related to this problem?
      ii. *If No (3 or below),* Why do you not think this indicates an issue with the {activity tracker}?
4. How would you react to this situation to be able to track your activity, fitness, or health?

<continue if the response does not include a why>

i. Prompt: What would you do and why?

ii. Prompt: what would make want to try that strategy for coping with this situation?

5. *If applicable,* Can you tell me what would you do if that did not work and the issue kept happening?

<if the response does not include a why>

i. Prompt: what would make want to try that/those/these strategy for coping with this situation?

C. Now imagine the very same situation. You have just walked all of Georgia Tech's Pi Mile trail, which you know to be exactly 3.14 miles. You notice that your {activity tracker} says you have walked 3.5 miles. However, you know this occurs every time you walk the Pi Mile trail.

3. What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale?
4. Can you please describe what you think is going on in this scenario?
5. Do you think this indicates an issue with the {activity tracker}?
    a. How confident are you in your judgment of this situation (that is whether or not this is an issue). Where 1=extremely confident it is not an issue and 6=extremely confident it is an issue.
        i. *If yes (4 or greater)*, what do you think the issue is?
            a. *If error that is specific or has a cause*, what would suggest that was the root of this issue?
                i. Prompt: As is in what signs are there that {insert what participant says} is related to this problem?
        ii. *If No (3 or below)*, Why do you not think this indicates an issue with the {activity tracker}?
6. How would you react to this situation to be able to track your activity, fitness, or health?

<continue if the response does not include a why>

i. Prompt: What would you do and why?

ii. Prompt: what would make want to try that strategy for coping with this situation?

7. *If applicable*, Can you tell me what would you do if that did not work and the issue kept happening?

    <if the response does not include a why>

i. Prompt: what would make want to try that/those/these strategy for coping with this situation?

D. Imagine you are reviewing your sleep data and find that your {activity tracker} says you only awoke once throughout the night. However, you recall waking up several times.

3. What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale?
4. Can you please describe what you think is going on in this scenario?
5. Do you think this indicates an issue with the {activity tracker}?
   a. How confident are you in your judgment of this situation (that is whether or not this is an issue). Where 1=extremely confident it is not an issue and 6=extremely confident it is an issue.
      i. *If yes (4 or greater)*, what do you think the issue is?
         a. *If error that is specific or has a cause,* what would suggest that was the root of this issue?
            i. Prompt: As is in what signs are there that {insert what participant says} is related to this problem?
      ii. *If No (3 or below),* Why do you not think this indicates an issue with the {activity tracker}?
6. How would you react to this situation to be able to track your activity, fitness, or health?

<continue if the response does not include a why>

   i. Prompt: What would you do and why?
   ii. Prompt: what would make want to try that strategy for coping with this situation?
7. *If applicable*, Can you tell me what would you do if that did not work and the issue kept happening?

   <if the response does not include a why>

   i. Prompt: what would make want to try that/those/these strategy for coping with this situation?

E. Imagine you can not get the wearable tracking device to respond at all, even after charging it. When you press the button on the device itself, none of the displays (including any screens or lights) appear. *<if they have questions on this one, you can repeat the "it is not responding at all" part>*

1. What would be the impact of this situation on your ability to track your activity, fitness, or health using your activity tracker on a 1=no impact to 4=serious impact scale?
2. Can you please describe what you think is going on in this scenario?
3. Do you think this indicates an issue with the {activity tracker}?
    a. How confident are you in your judgment of this situation (that is whether or not this is an issue). Where 1=extremely confident it is not an issue and 6=extremely confident it is an issue.
        i. *If yes (4 or greater)*, what do you think the issue is?
            a. *If error that is specific or has a cause,* what would suggest that was the root of this issue?
                i. Prompt: As is in what signs are there that {insert what participant says} is related to this problem?
        ii. *If No (3 or below),* Why do you not think this indicates an issue with the {activity tracker}?
4. How would you react to this situation to be able to track your activity, fitness, or health?

<continue if the response does not include a why>

i. Prompt: What would you do and why?
ii. Prompt: what would make want to try that strategy for coping with this situation?

5. *If applicable*, Can you tell me what would you do if that did not work and the issue kept happening?

    <if the response does not include a why>

    i. Prompt: what would make want to try that/those/these strategy for coping with this situation?

PPT ID _____

APPENDIX E

ACTIVITY TRACKER EXPLANATION FORM

We are interested in how you represent your activity tracker system to yourself. Please answer the following questions.

1. What is your activity tracker?

Make____________________ Model___________________

2. List the different features of your activity tracker.

<turn over>

3. Imagine you are explaining how your activity tracker works to someone who has never heard of it. What would you tell them?

4. **Draw** or **diagram** how your activity tracker works. Try to capture how the features you listed above work together. Please label where necessary.

APPENDIX F

CODING SCHEME FOR ACTIVITY TRACKER EXPLANATION FORMS

For accuracy: only add a point for non-repetitive wrong descriptions

1. List the different features of your activity tracker.

- One point for every feature or function participant mentions

2. Imagine you are explaining how your activity tracker works to someone who has never heard of it. What would you tell them?

- **Assign a point for key words:**
  - Detects or Senses
  - Calculates or Processes ---do not award points for this one for just "let's you know." Points for "let's you know" should occur with the information being shared. or looks for ("looks for repeated spikes")
  - figures out or know or tracks or records
  - Accelerometer or motion sensor or motion or walk or steps or pedometer*
  - Altimeter, height, or stair sensor or measures stairs*
  - Award extra points for specific measurement details (e.g., 10 ft = 1 flight of stairs, 2 steps = 1 count of step)
  - Explanations of why features work "Since people tend to sway their arms as they walk, tracking arm motion through the day should also give an accurate measure of how many steps have been taken."
  - Synch, blue tooth, or through a connection
  - Talks to or communicates with
  - Wearable or wear the device is worn
  - Where it should be worn (e.g., wrist)
  - Step size, height, or leg length
  - Calibrate
  - Calories
  - Nutrition (or specifics like salt, fats, vitamins, etc.)
  - Sensitivity setting (or similar names)
  - GPS
  - Pace
  - Weight
  - Distance
  - Manual input (on phone or website; 1 point total)
  - Custom logs
  - Visual summary (e.g., flower on fitbit, lights for progress, etc0
  - Measures/Tracks/Monitors/ Counting sleep*
  - Measures/Tracks/Monitors/ Counting distance/miles/how far/odometer
  - Measures/Tracks/Monitors/Counting _{other than steps, stairs, sleep, distance}___ You can include the flower on Fitbit here.
  - Other

- * a lot an extra point for every detail the tracker measures . This might mean multiple points per feature if different functions are mentioned. For example, *"it tracks your sleep and like when you wake up or your light sleep versus your deep sleep and stuff like that."* Would count for two points—1 point for the tracking when your asleep and one for tracking light vs deep sleep.
- **This is not an exhaustive list, if an additional facet of info, consider it an other**

3. **For the Draw or diagram how your activity tracker works… use this same coding scheme for captions plus award extra points for directional in drawings (e.g., where to wear the tracker, an arrow from the tracker to the phone)**

Award points for interactions

Make a list for each participant to justify interactions

Award points per each display

Make a list for each participant to justify display

PPT ID _____

APPENDIX G

DEVICE MENTAL MODEL QUESTIONNAIRE

Device Mental Model Knowledge Questionnaire

This questionnaire is designed to assess your knowledge of your activity tracker.

What is your activity tracker?

Make____________ Model________________

1. My activity tracker automatically tracks the flights of stairs climbed.	True	False
2. My activity tracker can or does use Bluetooth.	True	False
3. To make my activity tracker log my sleep, I can press a button on the wearable tracking device.	True	False
4. I can change the time for the alarm of my wearable tracking device from the wearable tracking device itself.	True	False
5. My activity tracker uses an altimeter[1].	True	False
6. My activity does *not* use an accelerometer[2].	True	False
7. I can modify some automatically logged data by using my activity tracker's phone app.	True	False
8. My activity tracker uses calories consumed to calculate my stride length.	True	False
9. Because it does not appear on any display, my stride length is *not* calculated or assumed by my activity tracker.	True	False
10. If my activity tracker is in sleep-mode and I move the part of my body where my device is worn (e.g., arm), poor sleep will likely be measured.	True	False

[1] altimeter: "an instrument for measuring altitude; *especially* **:** an aneroid barometer designed to register changes in atmospheric pressure accompanying changes in altitude" (Merriam-Webster)
[2] accelerometer: "an instrument for measuring acceleration or for detecting and measuring vibrations" (Merriam-Webster)

PPT ID _____

APPENDIX H

TECHNOLOGY EXPERIENCE PROFILE

Technology Experience Profile

The purpose of this set of questions is to assess your familiarity and experience with technology.

The following pages list technologies from different areas.

Please circle the most appropriate response to indicate how much you have used the technology listed, within the last 12 months.

PPT ID _____

Within the last 12 months, how much you have used…?					
Communication Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
1. Answering Machine/ Voicemail (e.g., record and retrieve messages) *with or without video relay service	1	2	3	4	5
2. Automated Telephone Menu System (e.g., pay bills, refill prescriptions) *with or without video relay service	1	2	3	4	5
3. Fax (e.g., receive and send printed documents)	1	2	3	4	5
4. Mobile Phone (e.g., make and receive calls) *with or without video relay service	1	2	3	4	5
5. Text Messaging (e.g., phone texting, BBM, iMessage, SMS)	1	2	3	4	5
6. Video call/conferencing (e.g., Skype, Facetime)	1	2	3	4	5
Within the last 12 months, how much you have used…?					
Computer Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
7. Desktop/Laptop Computer	1	2	3	4	5
8. Tablet Computer (e.g., iPad, Surface)	1	2	3	4	5
9. Email (e.g., Gmail, Yahoo)	1	2	3	4	5
10. Photo/Video Software (e.g., editing, organizing; iPhoto, Picture Manager, Photoshop)	1	2	3	4	5
11. Productivity Software (e.g., Excel, PowerPoint, Quicken, TurboTax, Word)	1	2	3	4	5

12. Social Networking (e.g., Facebook, MySpace)	1	2	3	4	5
Within the last 12 months, how much you have used…?					
Everyday Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
13. Automatic Teller Machine (ATM)	1	2	3	4	5
14. Photocopier (e.g., Lexmark, Xerox)	1	2	3	4	5
15. Home Security System (e.g., Ackerman Security, ADT)	1	2	3	4	5
16. In-Store Kiosk (e.g., grocery self-checkout, price checker)	1	2	3	4	5
17. Microwave Oven	1	2	3	4	5
18. Programmable Device (e.g., coffee maker, thermostat)	1	2	3	4	5
Within the last 12 months, how much you have used…?					
Health Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
19. Blood Pressure Monitor (e.g., measure blood pressure)	1	2	3	4	5
20. Digital Thermometer (e.g., measure temperature)	1	2	3	4	5
21. Health Management Software (e.g., to keep track of weight, diet, exercise; Personal Health Record)	1	2	3	4	5
22. Heart Rate Monitor (e.g., measure heart rate, pulse)	1	2	3	4	5
23. Medication Reminder Device (e.g., schedule electronic alerts)	1	2	3	4	5
24. Pedometer (e.g., measure walking distance)	1	2	3	4	5

Within the last 12 months, how much you have used…?					
Recreational Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
25. Digital Music Player (e.g., iPod, MP3 player, Zune, tablet)	1	2	3	4	5
26. Digital Photography (e.g., camera, tablet, phone)	1	2	3	4	5
27. Electronic Book Reader (e.g., Kindle, Nook, tablet)	1	2	3	4	5
28. Gaming Console (e.g., Playstation, Wii, XBox)	1	2	3	4	5
29. Online Shopping/Coupons (e.g., Amazon, Groupon, retail stores)	1	2	3	4	5
30. Recording and Playback Device (e.g., Blu-Ray, CD, DVD, DVR, VCR)	1	2	3	4	5
Within the last 12 months, how much you have used…?					
Transportation Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
31. Airline Kiosk (e.g., check in, print boarding pass)	1	2	3	4	5
32. Bus Tracker (e.g., check location of buses, estimate time of arrival)	1	2	3	4	5
33. Online Map Software (e.g., get directions, plan routes; Google Maps, MapQuest)	1	2	3	4	5
34. In-Vehicle Navigation System/GPS	1	2	3	4	5
35. Online Travel Reservation (e.g., airline website, Expedia, Travelocity)	1	2	3	4	5
36. Parking Payment System (e.g., exiting lot, paying for space)	1	2	3	4	5

Within the last 12 months, how much you have used…?					
Other Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
37. Blood Glucose Meter Control (e.g., controls glucose levels)	1	2	3	4	5
38. Hearing Aids (e.g., helps with hearing)	1	2	3	4	5
39. Car Seat Adjustment (e.g., moves seat)	1	2	3	4	5
40. Washing/and or Drier (e.g., laundry machines)	1	2	3	4	5
41. Alarm Clock (e.g., a bell or buzzer set to a time)	1	2	3	4	5
42. Iron (e.g., for clothes or linens)	1	2	3	4	5
43. Cruise Control (e.g., in a car)	1	2	3	4	5
44. Remote Control (e.g., for television, stereo, DVD player)	1	2	3	4	5
45. Scale (e.g., bathroom scale, a balance)	1	2	3	4	5
46. Vacuum (e.g., a dirt devil)	1	2	3	4	5

Within the last 12 months, how much you have used…?					
Wellness Management Technology	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
47. Any wearable activity tracker (e.g., Fitbit, Jawbone Up, Nike Fuelband)	1	2	3	4	5
48. A log for calorie consumption (e.g., Myfitnesspal, Loseit!)	1	2	3	4	5
49. A separate wellness	1	2	3	4	5

management technology **with** your current activity tracker (e.g., linking Jawbone to Myfitnesspal)					

In considering only your current activity tracker and its corresponding website and phone app, please answer how frequently you've used the following features:

Within the last 12 months, how much you have used, in any way, ...?					
Current Activity Tracker	Not sure what it is	Not used	Used once	Used occasionally	Used frequently
50. The automated wearable tracking device (e.g., the Jawbone Up bracelet)	1	2	3	4	5
51. The website	1	2	3	4	5
52. The phone app	1	2	3	4	5
53. The step tracker (e.g., steps walked/ran)	1	2	3	4	5
54. The calorie tracker	1	2	3	4	5
55. The distance tracker	1	2	3	4	5
56. The sleep tracker	1	2	3	4	5
57. The alarm	1	2	3	4	5
58. The manual data entry for the food log	1	2	3	4	5
59. The manual data entry for the exercise log	1	2	3	4	5
60. The manual data entry for the sleep log	1	2	3	4	5
61. The share data with friends feature (e.g., teams, competitions, challenges)	1	2	3	4	5
62. Other feature or log (please list) _______	1	2	3	4	5
63. Other feature or log (please	1	2	3	4	5

list) _______					
64. Other feature or log (please list) _______	1	2	3	4	5
65. Other feature or log (please list) _______	1	2	3	4	5
66. Other feature or log (please list) _______	1	2	3	4	5
67. Other feature or log (please list) _______	1	2	3	4	5
68. Other feature or log (please list) _______	1	2	3	4	5
69. Other feature or log (please list) _______	1	2	3	4	5
70. Other feature or log (please list) _______	1	2	3	4	5

APPENDIX I

AUTOMATION-INDUCED COMPLACENCY POTENTIAL QUESTIONNAIRE

PPT ID _____

Automation Experience Questionnaire

Read each statement carefully and circle the one response that you feel most accurately describes your views and experiences. THERE ARE NO RIGHT OR WRONG ANSWERS. Please answer honestly and do not skip any questions.

SD	D	U	A	SA
Strongly disagree	Disagree	Undecided	Agree	Strongly agree

1. Manually sorting through emails is more reliable than computer-aided searches for finding emails in my inbox.	SD D U A SA
2. If I need to have a tumor in my body removed, I would choose to undergo computer-aided surgery using laser technology because computerized surgery is more reliable and safer than manual surgery.	SD D U A SA
3. People save time by using automatic teller machines (ATMs) rather than a bank teller in making transactions.	SD D U A SA
4. I do not trust automated devices such as ATMs and computerized pay stations for parking lots.	SD D U A SA
5. People who work frequently with automated devices have lower job satisfaction because they feel less involved in their job than those who work manually.	SD D U A SA
6. I feel safer depositing my money at an ATM than with a human teller.	SD D U A SA
7. I have to pay an important bill. To ensure that the bill is paid with the correct amount and on time, I would use the automatic bill pay facility on my online banking rather than pay the bill manually.	SD D U A SA
8. People whose jobs require them to work with automated systems are lonelier than people who do not work with such devices.	SD D U A SA
9. Automated systems used in modern aircraft, such as the automatic landing system, have made air journey safer.	SD D U A SA
10. ATMs provide safeguard against the inappropriate use of an individual’s bank account by dishonest people.	SD D U A SA
11. Automated devices used in aviation and banking have made work easier for both employees and customers.	SD D U A SA
12. I often use automated devices.	SD D U A SA
13. People who work with automated devices have greater job satisfaction because they feel more involved than those who work manually.	SD D U A SA
14. Automated devices in medicine save time and money in the diagnosis and treatment of disease.	SD D U A SA

15.	Even though the automatic cruise control in my car is set at a speed below the speed limit, I worry when I pass a police radar speed-trap in case the automatic control is not working properly.	SD D U A SA
16.	Bank transactions have become safer with the introduction of computer technology for the direct deposit of checks.	SD D U A SA
17.	I would rather purchase an item using a computer than have to deal with a sales representative on the phone because my order is more likely to be correct using the computer.	SD D U A SA
18.	Work has become more difficult with the increase of automation in aviation and banking.	SD D U A SA
19.	I do not like to use ATMs because I feel that they are sometimes unreliable.	SD D U A SA
20.	I think that automated devices used in medicine, such as CAT-scans and ultrasound, provide very reliable medical diagnosis.	SD D U A SA

APPENDIX J

CHAMPS QUESTIONNAIRE

CHAMPS Activities Questionnaire

The purpose of this questionnaire is to assess the physical activities you take part in on a weekly bases. It asks about activities that you may have done in the past 4 weeks. The questions on the following pages are similar to the example shown below.

INSTRUCTIONS

1.1 If you DID the activity in the past 4 weeks:

Step #1 **Check the YES box.**

Step #2 **Think about how many TIMES a week you usually did it, and write your response in the space provided.**

Step #3 **Circle how many TOTAL HOURS in a typical week you did the activity.**

Here is an example of how Mrs. Jones would answer question #1: Mrs. Jones usually visits her friends Maria and Olga twice a week. She usually spends one hour on Monday with Maria and two hours on Wednesday with Olga. Therefore, the total hours a week that she visits with friends is 3 hours a week.

In a typical week during the past 4 weeks, did you…							
1. Visit with friends or family (other than those you live with)? ☒ **YES** How many TIMES a week? 2 ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

1.1 If you DID NOT do the activity:

- **Check the NO box and move to the next question**

In a typical week during the past 4 weeks, did you …								
1. Do woodworking, needlework, drawing, sketching or other arts or crafts? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
2. Attend a concert or sport event? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
3. Play cards, board games, billiards, or a musical instrument? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
4. Do work around the house (such as washing windows, sweeping, vacuuming)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
5. Use a rowing machine or row? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
6. Go for a jog, a run, or a sprint session <u>on a treadmill</u>? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
7. Go for a jog, run, or sprint <u>session outside or on a track</u>? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
8. Walk uphill or hike uphill (count only uphill part) <u>on a treadmill</u>? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
9. Walk uphill or hike uphill (count only uphill part) <u>outside</u>? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
10. Walk <u>leisurely</u> to do errands, to attend classes or meetings, to exercise, or for pleasure? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
11. Walk <u>fast or briskly</u> for exercise (do <u>not</u> count walking leisurely or uphill, include treadmill)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
12. Use an elliptical? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
13. Use a stair master, stair machine, or exercise on stairs? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
14. Ride a bicycle or stationary cycle? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
15. Do yoga or Tai-chi? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
16. Do aerobics or aerobic dancing? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL <u>hours a week</u> did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
17. Do strength-training exercises (such as hand-held weights, weight machines, or push-ups)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
18. Play basketball, soccer, racquetball, tennis, baseball, football, rugby or other team sports (do not count time on sidelines)? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
19. Do water exercises or swim? ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours
20. Take any exercise classes (e.g., zumba, spin)? ☐ **YES** **What class? __________** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

In a typical week during the past 4 weeks, did you …							
21. Do other types of physical activity not previously mentioned (please specify)? ________________________ ☐ **YES** How many TIMES a week?_____ ➔ ☐ **NO**	How many TOTAL hours a week did you usually do it? ➔	Less than 1 hour	1-2½ hours	3-4½ hours	5-6½ hours	7-8½ hours	9 or more hours

APPENDIX K

BACKGROUND AND HEALTH QUESTIONNAIRE

BACKGROUND QUESTIONNAIRE

For HFA Personnel Use Only DATE: __ __ / __ __ / __ __

SUBJECT ID: __ __ __ - __ __ __ __ -- __ __ __ __ __ -- __ __

Data Entered By: DATE

1st ______________________________ __ __ / __ __ / __ __

2nd ______________________________ __ __ / __ __ / __ __

☐ Consent given to include data in archived repository

Thank you for participating on our research!

This questionnaire asks you to provide information about various aspects of your background, including your demographic and health information. Please answer the questions by placing an X in the appropriate box.

Published documents regarding these answers will not identify individuals with their answers. However, if there is a question that you do not wish to answer, please leave it blank and go on to the next question.

Demographic Information

1. Gender: ☐1 Male ☐2 Female

2. What is your date of birth? _________________________
(mm/dd/yyyy)

3. Are you fluent in English? ☐1 Yes ☐2 No

4. What is your preferred language for communicating?
   - ☐1 English
   - ☐2 Spanish
   - ☐3 American Sign Language
   - ☐4 Other (please list)___________________

5. What is your highest level of education?
   - ☐1 No formal education
   - ☐2 Less than high school graduate
   - ☐3 High school graduate/GED
   - ☐4 Vocational training
   - ☐5 Some or in-progress college/Associate's degree
   - ☐6 Bachelor's degree (BA, BS)
   - ☐7 Master's degree (or other post-graduate training)
   - ☐8 Doctoral degree (PhD, MD, EdD, DDS, JD, etc)
   - ☐9 Do not wish to answer

6. Current marital status (Check **one**)
   - ☐1 Single
   - ☐2 Married
   - ☐3 Separated
   - ☐4 Divorced
   - ☐5 Widowed
   - ☐6 Other (please specify) __________________
   - ☐7 Do not wish to answer

7. Do you consider yourself Hispanic or Latino?

☐1 Yes ☐2 No ☐3 Do not wish to answer

8. How would you describe your primary racial group?
   - ☐1 American Indian/Alaska Native
   - ☐2 Asian
   - ☐3 Black or African American
   - ☐4 Native Hawaiian or Other Pacific Islander
   - ☐5 White
   - ☐6 More than one race
   - ☐7 Other (please specify) ________________________
   - ☐8 Do not wish to answer

9. In which type of housing do you live?
   - ☐1 Single family home
   - ☐2 Apartment or Condominium
   - ☐3 Assisted living residence
   - ☐4 Nursing home residence
   - ☐5 Other (please specify) __________________
   - ☐6 Do not wish to answer

10. Which one of the following BEST describes your living arrangement?
   - ☐1 Living alone
   - ☐2 Living with your immediate family (i.e., spouse/partner and/or dependent children, or parents if never married)
   - ☐3 Living with your adult children
   - ☐4 Living with your (or your spouse/partner's) extended family (e.g., parents, siblings, cousins)
   - ☐5 Living with roommate(s)
   - ☐6 Other (please specify) __________________
   - ☐7 Do not wish to answer

11. Is your housing or community specifically designed for seniors (i.e., 55 and older)?

☐1 Yes ☐2 No ☐3 Not sure

12. What is your primary mode of transportation? (Check **one**)

- ☐1 Drive myself
- ☐2 A friend or family member drives me
- ☐3 Walk
- ☐4 Bicycle
- ☐5 Taxi
- ☐6 Use transportation service provided by my residence
- ☐7 Use public transportation (e.g., bus, subway, van services)
- ☐8 Other (please specify) ___________________

13. Which category best describes your yearly household income? Do not give the dollar amount, just check the category.

- ☐1 Less than $25,000
- ☐2 $25,000 - $49,999
- ☐3 $50,000 - $74,999
- ☐4 $75,000 or more
- ☐5 Do not wish to answer
- ☐6 Do not know for certain

Occupational Status

14a. What is your primary occupational status? (Check **one**)

- ☐1 Employed full-time Occupation? ___________________________________
- ☐2 Employed part-time Occupation? ___________________________________
- ☐3 Student
- ☐4 Homemaker
- ☐5 Retired Former occupation? ____________________ Year retired? __________
- ☐6 On maternity leave, on sick leave, or on disability benefits
- ☐7 Unemployed or temporarily laid off
- ☐8 Other (please specify) ____

14 b. If applicable, what was or is your college major? If undecided, please write undecided with your college (e.g., undecided, College of Sciences).

Other

1. What is your height?

___________Feet ________Inches

Health Information

1. In general, would you say your health is:

☐1 ☐2 ☐3 ☐4 ☐5

Poor Fair Good Very good Excellent

2. Compared to other people your own age, would you say your health is:

☐1 ☐2 ☐3 ☐4 ☐5

Poor Fair Good Very good Excellent

3. How satisfied are you with your present health?

☐1 ☐2 ☐3 ☐4 ☐5

Not at all satisfied Not very satisfied Neither satisfied nor dissatisfied Somewhat satisfied Extremely satisfied

4. How often do health problems stand in the way of your doing the things you want to do?

☐1 ☐2 ☐3 ☐4 ☐5

Never Seldom Sometimes Often Always

5. How many different **prescription medications** do you take <u>each day</u>?

____

6. How many different **over-the-counter medications/supplements** do you take <u>each day</u>?

____

7. Please indicate if you have ever been told by a health professional that you have any of the following conditions. Check **<u>one</u>** box for each condition.

Condition	**Yes₁**	**No₂**	**Do not wish to answer/ Not sure₃**
a. Alzheimer's Disease			
b. Arthritis			
c. Asthma			
d. Cancer			
e. Cardiac Atrial Fibrillation/ Cardiac Arrhythmia			
f. Chronic Kidney Disease			
g. Chronic Obstructive Pulmonary Disease (COPD)			
h. Coronary Artery Disease/ Coronary Heart Disease			
i. Depression			
j. Diabetes/High Blood Sugar			
k. Heart Failure/ Congestive Heart Failure			
l. High Blood Pressure/Hypertension			
m. High Cholesterol/Hyperlipidemia			
n. Osteoporosis			
o. Overweight			
p. Stroke/Transient Ischemic Attack			
q. Other? (If yes, please list below)			

Condition	Yes1	No2	Do not wish to answer/ Not sure3
______________________________ ______________________________			

8. If you are on a diet, please describe it here. Otherwise, write not applicable:

____

____

____

____

____

____

____
_

9. Please list any physical activity you engage in on a weekly basis because of health advice (e.g., doctor's orders, physical therapy). Otherwise, write not applicable:

____

____
_

APPENDIX L

AUTOMATION MEASUREMENT PROFILE

Automation Measurement Profile

This questionnaire is designed to assess your perceptions of your activity tracker's accuracy. For each prompt below, imagine that your activity tracker has stated the given measurement. Please place an X in the box(es) you believe you may have actually walked, slept, moved, or burnt. Please note you are NOT limited to one box per prompt.

Your activity tracker reports you have walked 7,000 steps.	5,950 steps	6,300 steps	6,650 steps	7,000 steps	7,350 steps	7,700 steps	8,050 steps
How many steps do you think you have walked?							
Your activity tracker reports you have slept for 8 hours.	6 hrs 48 min	7 hrs 12 min	7hrs 36 min	8hrs 0 min	8 hrs 24 min	8 hrs 48 min	9 hrs 12 min
How many long do you think you have slept?							
Your activity tracker reports you have moved (walked/ran) 4 miles.	3.4 miles	3.6 miles	3.8 miles	4 miles	4.2 miles	4.4 miles	4.6 miles
How far do you think you have moved?							

Your activity tracker reports you have burnt 1,500 calories today.	1275 cal.	1350 cal.	1425 cal.	1,500 cal.	1575 cal.	1650 cal.	1725 cal.
How many calories do you think you have burnt?							

APPENDIX M

EXPERIMENT SCRIPT

Experiment Script

*For all questionnaires, read the directions out loud with the participant.

Before the participant arrives, have the informed consent, questionnaires, interview script, and recorder ready.

Hello, <Participant Name>, how are you today? Thank you for being part of our study. We couldn't do this without your help. This study involves questionnaires and interviews that assess your experience with either the Fitbit One or the Jawbone Up 24 or {name of their activity tracker}. Technologies like the Fitbit One or the Jawbone Up 24 or{name of their activity tracker}. use automation to help people track their physical activity and health. We are particularly interested in the sorts of errors you may come across in using the automation on the Fitbit One or the Jawbone Up 24 or {name of their activity tracker}. Before we continue, would you please silence your cell phone? <wait for participant to silence cell phone> Thank you. Now, I'd like to go through the study in more detail with you, and get your consent to take part in the study. This is our informed consent form. We'll go over it together now and please ask me any questions you have about it.

Go through informed consent (2 copies). Make sure everyone signs where needed before continuing.

Before we begin, can I please see your Fitbit One or Jawbone Up 24 or {other activity tracker} to check what it is?

Check that they have the correct technology. Note: Jawbone says what model it is inside the wrist band.

Great! So anytime you hear "your activity tracker" today, we'll be talking specifically about your {activity tracker name}. I now have a few more questions about your experience with Fitbit One or Jawbone Up 24 or {activity tracker} to make sure you qualify for the study and to get more information on your usage background.

Give participant Wellness Management Technology Experience Background Questionnaire. Check for completion. Check that participant uses the technology 3 days a week for at least 1 month. If he or she does not meet this usage requirement, dismiss the participant: I'm sorry, but unfortunately, you do not have the usage experience we require for this study, so we will not be able to continue. Thank you for your time and interest. Dismiss participant.

If participant meets usage requirements: Great, you do have the experience we're looking for in this study. So we will now continue. Please stop me at any time if you have any questions or need a break. The next part of our study is an interview. This interview will take an hour to complete, so please feel free to take a break now to get water or use the restroom. If you need to take a break during the interview, please let me know and I will pause the interview. Please answer all questions freely and honestly. Also, some of these questions may seem repetitive. It's okay if your answers overlap or are different to them, I just want to make sure I'm asking the same questions to everyone. Even if it feels like you're repeating yourself, please answer those questions so that we can make sure we're capturing your responses with where we'll be looking for them. I'm going to turn on the recorder now.

Turn on recorder.

Interview Script.

Turn recorder off.

Thank you. I have now turned off the recorder. That is the end of the interview portion of this study, but we have several questionnaires left for you to complete. These should take about half an hour to complete. Before doing so, we'll take a 5 minute break. Please feel free to use the restroom or get some water during this time.

Alright, let's continue. For this part, we will ask to explain how your {activity tracker} works and draw a flow chart or diagram of how you think the your {activity tracker}. <If participant asks to use their app or device while completing this, please ask them not to.>

Fitbit One/Jawbone Up/Activity Tracker Explanation Form.

For this next questionnaire, circle if the statements are true or false for {your Fitbit or Jawbone or activity tracker}. If you don't know one of the questions, please take your best guess. <If participant asks to use their app or device while completing this, please ask them not to.>

Device Model Knowledge Questionnaire. Check for completion.

Next, we have some questions about your technology experience. We will also ask you to draw a flow chart of diagram with how you think {your Fitbit One or Jawbone Up 24 or your activity tracker} works.

Technology Experience Profile. Check for completion.

Thank you. Next, I have some questions about your experiences and thoughts on automation.

Automation Experience Profile. Check for completion.

Thank you. Now I have some questions about the sorts of physical activities you typically take part in.

CHAMPS Questionnaire. Spend extra time on the example, asking if the format for answering makes sense, and re-explain of it does not. Check for completion.

Thanks! We only have two more questionnaires left. For this next one, we just want to know a little about your background and health.

Demographics and Health Questionnaire. Check for completion.

Thank you. For our last one, we want to know about how accurate you think the {Fitbit One/Jawbone Up 24/Activity tracker} is.

Automation Measurement Profile. Check for completion.

Thank you, <Participant Name> for your time! We could not have done this without your help. If you have any questions for us, please let us know, and we are finished with the study! Now we will debrief you on the research study you just participated in.

Debriefing form. Give a copy to the participant. Explain the purpose and goals for the experiment, answer any questions they may have, and thank them for their time. Then, after the participant leaves, add credit to SONA. OR have participant sign for receipt of check and provide them a check.

Appendix N

Strategies That Emerged from the TSI and the SBI Strategy Coding Scheme

Ref #	Strategy Group	Strategy
301	Change or Monitor Your Behavior in the Situation	carry out certain movements or stay still to make tracker more accurate *"Moving your arm to match your step pattern."*
302	Change Usage Pattern	Temporarily stop using activity tracker *"turn it off before I got on the horse" "or I could even take the Up off my wrist and put it in the console of my car"*
303	Try to Fix it on Your Own	restart/reset/delete and reinstall device, app, phone, or computer, potentially resynch thereafter *"I'd probably try to reset."*
304	Change or Monitor Your Behavior in the Situation	check for human error *"Maybe ensuring that I was following the race course, I'd say"*
305	Change Usage Pattern	synch at a different time/in a different place *"Sometimes I try and sync it right after I've worked out to see if that will help it sync."*
306	Continue to Use	estimate activity/calculate in your mind/correct in your mind *"just did the mental understanding it was going to be overestimated and just reversing it in my mind."*
307	Try to Fix it on Your Own	replace part of activity tracker *"to get a new watch band that has an adjustable strap"*
308	Gather Information or Seek Help to Get it Fixed	get help online *"I look up the frequently asked questions on the Fitbit website and I follow that"*
309	Change Usage Pattern	only use activity tracker for a specific activity *"maybe only wear it when I'm working out I guess."*
310	Change Usage Pattern	Stop using feature forever *"my other alternative is to just turn off all the alarms" "Not to use that feature"*
311	Change Usage Pattern	Stop using whole tracker forever *"the best way to do it would probably just not to use the….device itself" "I would stop doing*

		it.”
312	Change or Monitor Your Behavior in the Situation	Do something/change your behavior to prevent error *“so I would try and keep it from getting wet” “so I get the towel and try to like wipe off my arm before I do it and just keep it looser to where when I get home”*
313	Change or Monitor Your Behavior in the Situation	Change your behavior (do or do not do something) to work around the error (not to prevent) *“been checking the app more often to see how much battery there is left” “I just generally tried to eat healthier, which is a little, a bit hard to describe, but you just you know, eat a salad for lunch instead. Which is a little bit more intangible”*
314	Try to Fix it on Your Own	try to synch with something else *“we don’t sync on each other’s computer anymore”*
315	Wait for Something to Happen	wait for feedback or warning from the tracker *“wait and see if my tracker either gives me the low battery warning for me to plug it in”*
316	Gather Information or Seek Help to Get it Fixed	troubleshoot/test the tracker *“I guess in the same way that I’ve tried to see how steps are being counted I could actually start from like zero flights and actually go up a bunch of steps and see how it actually is.”*
317	Try to Fix it on Your Own	use rice (un-do the cause) “*Putting it in rice as soon as possible”*
318	Wait for Something to Happen	wait it out “*I’ll just wait it out...Just I’ll not sync it, and hopefully it will sync another time.”*
319	Change Usage Pattern	use a method other than activity tracking (e.g., buying pre-made meals) *“there’s also all of the meals that you preorder online that’s like here’s your calorie half of the day in these tiny little pre-heatable boxes too. So that would be an alternative to just tracking the food that you make yourself of just being given a regimented diet.”*
320	Change Usage Pattern	manually entering data into website or app “*enter in the information yourself”*
321	Continue to Use	keep using device despite error/ignore error *“I would just brush it off as a one time weird occurrence” “I could just turn off the sleep mode and not worry about the exercise time lost, and that data that was lost as well.”*
322	Change Usage Pattern	manually keep track on paper or excel “*Manually write it down”*
323	Gather Information or Seek Help to Get it Fixed	monitor for future errors “*Currently I now touch down and make sure all my, my Fitbit is synced and that the alarms are at the correct time”*

324	Change Usage Pattern	use a different tracker or app as primary for at least one feature "*I mean there's also a step counter that comes with the iPhone 6… I could always check that."*
325	Change Usage Pattern	use a different tracker, app, technology as a back up method *"or rely on a phone alarm instead of the Fitbit alarm."*
326	Change Usage Pattern	synch other apps to tracker (so use a different tracker/app that feeds into main activity tracker) "*trying to lose weight I will use… So I sync the MyFitnessPal with the [Activity Tracker] app so I get the step count, and I don't correct that"*
327	Change Usage Pattern	return/replace/send to manufacturer *"I probably would buy a new device." "I'd send it back, because it is affecting everything"*
328	Try to Fix it on Your Own	repeat earlier strategies *"probably try some other things, or try whatever I tried before a few more times"*
329	Change Usage Pattern	augment tracker with manual records that are informed by online health database *"so augment it with a physical journal informed by whatever the database online was, that would tell me more information about it."*
330	Gather Information or Seek Help to Get it Fixed	bring to a place, other than the company, that will fix it *"tracker service stations. I mean if there's some place that could fix it I would probably take it there."*
331	Gather Information or Seek Help to Get it Fixed	Research the tracker *"maybe try to do some research to try to figure out, what how it counts to floors"*
332	Change or Monitor Your Behavior in the Situation	adjust how/where you wear tracker *"Probably just repositioning it"*
333	Try to Fix it on Your Own	adjust the sensitivity of the tracker *"you can decrease the sensitivity of the pedometer"*
334	Change or Monitor Your Behavior in the Situation	be more active *"… I maybe have been a little more active in terms of if I have the option to say sit and watch TV or clean up the house or the apartment for example, then that will at least have me moving"*
335	Try to Fix it on Your Own	scroll through wearable device screens *"sometimes I'll scroll through the screen on my fitbit one(wearable portion) to wake it up a little bit"*
336	Try to Fix it on	software update or anything about hardware/software mismatch

	Your Own	*"updating the [ACTIVITY TRACKER/ACTIVITY TRACKER BRAND] software on my phone and computer"*
337		contact company *"Contact the manufacturer." "I could have complained" "I emailed Fitbit customer support, and ultimately they wound up sending me a new one"*
439*		Other- a strategy not covered by the above
100		Null – no action/strategy is provided, only cues

*Code 439 was added to capture any "other" strategies mentioned in the SBI that might not have been mentioned in the TSI.

REFERENCES

Andersson, B. E., & Nilsson, S. G. (1964). Studies in the reliability and validity of the critical incident technique. *Journal of Applied Psychology*, *48*(6), 398-403.

Bagheri, N., & Jamieson, G. A. (2004, October). The impact of context-related reliability on automation failure detection and scanning behaviour. In IEEE Proceedings *Systems, Man and Cybernetics,* 1, 212-217.

Barg-Walkow, L. H. (2013). *Understanding the role of expectations of human responses to an automated system.* (Unpublished master's thesis). The Georgia Institute of Technology, Atlanta.

Barg-Walkow, L. H., Mitzner, T. L., & Rogers, W. A. (2014*). Technology Experience Profile* (HFA-TR-1402). Atlanta, GA: Georgia Institute of Technology, School of Psychology, Human Factors and Aging Laboratory.

Bereiter, S. R., & Miller, S. M. (1989). A field-based study of troubleshooting in computer-controlled manufacturing systems. In IEEE Proceedings Transactions on *Systems, Man and Cybernetics 19*(2), 205-219.

Bisantz, A. M., & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics*, *28*(2), 85-97.

Bodenheimer, T., Lorig, K., Holman, H., & Grumbach, K. (2002). Patient self-management of chronic disease in primary care. *JAMA: The Journal of the American Medical Association*, *288*(19), 2469-2475.

Bujnoch, Z. (2011). Home healthcare and disease management for remote monitoring market: A cornerstone to tele-health. San Antonio, TX: Frost & Sullivan Research Services.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology,74*(3), 271-280.

Consolvo, S., McDonald, D. W., & Landay, J. A. (2009, April). Theory-driven design strategies for technologies that support behavior change in everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 405-414). ACM.

Consolvo, S., McDonald, D. W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., ... & Smith, I. (2008, April). Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1797-1806). ACM.

Craik, K. (1943). *The Nature of Explanation.* Cambridge, MA: Cambridge UP.

Czaja, S. J., Charness, N., Dijkstra, K., Fisk, A. D., Rogers, W. A., & Sharit, J. (2006). *Center for Research and Education on Aging and Technology Enhancement Demographic and Background Questionnaire* (CREATE-TR-2006-02). University of Miami: CREATE.

Craft, T. M., Parr, M. J., & Nolan, J. P. (2004). *Key topics in critical care*. Oxon, UK: CRC Press.

Business Insider (2014*). The wearable computing market report: growth trends, consumer attitudes, and why smartwatches will dominate*. New York, NY: T. Danova.

Durso, F. T., Ferguson, A. N., Kazi, S., Cunningham, C., & Ryan, C. (2015). Strategic threat management: An exploration of nursing strategies in the pediatric intensive care unit. *Applied Ergonomics*, *47*, 345-354.

Durso, F. T., Kazi, S., & Ferguson, A. N. (2014). The threat-strategy interview. *Applied Ergonomics*.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58,* 697-718.

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, *51*(4), 327-358.

Golding, S. L., & Rorer, L. G. (1972). Illusory correlation and subjective judgment. *Journal of Abnormal Psychology, 80*(3), 249-260.

Gregg, S. E., Ferguson, A. N., & Durso, F. T. (2014). A method for aggregating strategies across individuals managing a lack of IV access. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 644-648). SAGE Publications.

Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*(2), 336-355.

Holden, R. J., Rivera-Rodriguez, A. J., Faye, H., Scanlon, M. C., & Karsh, B. T. (2013). Automation and adaptation: nurses' problem-solving behavior following the implementation of bar-coded medication administration technology. *Cognition, Technology & Work*, *15*(3), 283-296.

Holtz, B., & Lauckner, C. (2012). Diabetes management via mobile phones: A systematic review. *Telemedicine and e-Health*, *18*(3), 175-184.

Itoh, M., Abe, G., & Tanaka, K. (1999). Trust in and use of automation: Their dependence on occurrence patterns of malfunctions. In IEEE Proceedings of *Systems, Man, and Cybernetics*, 3, 715-720.

Johnson, J. D. (2004). *Type of automation failure: The effects on trust and reliance in automation*. (Unpublished master's thesis). Georgia Institute of Technology, Atlanta, GA.

Lee, J. M., Kim, Y., & Welk, G. J. (2014). Validity of Consumer-Based Physical Activity Monitors. *Medicine and Science in Sports and Exercise*.

Lees, M. N., & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics*, *50*(8), 1264-1286.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50-80.

Kelley, H. H. (1967). Attribution theory in social psychology. In *Nebraska symposium on motivation*. University of Nebraska Press.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211.

Konradt, U. (1995). Strategies of failure diagnosis in computer-controlled manufacturing systems: empirical analysis and implications for the design of adaptive decision support systems. *International Journal of Human-Computer Studies*, *43*(4), 503-521.

Kuhn, T. (1996). *The structure of scientific revolutions.* Chicago: University of Chicago Press. (Original work published 1962)

Mackinlay, M. Z. (2013). Phases of accuracy diagnosis:(In) visibility of system status in the Fitbit. *Intersect: The Stanford Journal of Science, Technology and Society*, *6*(2).

Madhavan, P., Wiegmann, D. A., & Lacson, F. C. (2006). Automation failures on tasks easily performed by operators undermine trust in automated aids. *Human Factors, 48*(2), 241-256.

Masalonis, A. J. (2003, October). Effects of training operators on situation-specific automation reliability. In IEEE Proceedings of *Systems, Man and Cybernetics,* 3, 1595-1599).

McConnell, A. R., Sherman, S. J., & Hamilton, D. L. (1997). Target entitativity: implications for information processing about individual and group targets. *Journal of Personality and Social Psychology*, *72*(4), 750-762.

Moray, N., Inagaki, T., & Itoh, M. (2000). Adaptive atomation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, *6*, 44-58.

Muir, B. M., & Moray, N. (1996). Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, *39*(3), 429-460.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37*(11), 1905-1922.

Norman, D. A. (2002). *The design of everyday things*. New York, New York: Basic books.

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. (1975). A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology*, *32*(4), 605-616.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230-253.

Piaget, J. (1960). *The child's conception of the world*. London: Routleledge.

Pop, V. L. (2012). *Individual differences in the calibration of trust in automation*. (Unpublished master's thesis). Georgia Institute of Technology, Atlanta, GA.

Pop, V.L. & Stearman, E.J. (2015). An Updated Automation Induced Complacency Rated Scale [Measurement Instrument].

Purch (2014). Fitness Tracker Reviews. Retrieved from http://fitness-trackers-review.toptenreviews.com/

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman and M. Mouloua (Eds*.) Automation And Human Performance: Theory And Applications. Human Factors In Transportation*., (pp. 19-35). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Risen, J. L., & Gilovich, T. (2008). Why people are reluctant to tempt fate. *Journal of Personality and Social Psychology*, *95*(2), 293-307.

Rogers, E. M. (2003). *Diffusion of innovations* (5th edition). New York, NY: Free Press.

Rovira, E., Cross, A., Leitch, E., & Bonaceto, C. (2014). Displaying contextual information reduces the costs of imperfect decision automation in rapid retasking of isr assets. *Human Factors*, *56*(6), 1036-1049.

Sanchez, J. (2009,). Conceptual model of human-automation interaction. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting* (pp. 1403-1407). Santa Monica, CA: Human Factors and Ergonomics Society.

Sauro, J. (2013) Raiting the Severity of Usability Problems. Retrived from http://www.measuringu.com/blog/rating-severity.php

Singh, I. L., Molloy, R., & Parasuraman, R. (1993a). Automation-induced "complacency": Development of a complacency-potential scale. *International Journal of Aviation Psychology*, 3, 111-122.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, *96*(1), 58-.83

Stewart, A. L., Mills, K. M., King, A. C., Haskell, W. L., Gillis, D., & Ritter, P. L. (2001). CHAMPS physical activity questionnaire for older adults: outcomes for interventions. *Medicine & Science in Sports & Exercise, 33*(7), 1126-1141.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220-247.

Van der Veer, G. C., & Melguizo, P. MC (2003). Mental models. In A. Sears and J. A. Jacko (Ed.), *The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications* (pp. 55-80). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201-212.

Willaims, J. D. (1992). *Monitoring school performance: A guide for educators*. Washington, D.C.: The Falmer Press.

Woolsey, L. K. (1986). The critical incident technique: An innovative qualitative method of research. *Canadian Journal of Counselling, 20*(4), 242-254.